

Encyclopedia of
Complexity and Systems Science Series
Editor-in-Chief: Robert A. Meyers

SPRINGER NATURE []
Reference

Cesar E. Silva
Alexandre I. Danilenko
Editors

Ergodic Theory

A Volume in the Encyclopedia of
Complexity and Systems Science,
Second Edition

Encyclopedia of Complexity and Systems Science Series

Editor-in-Chief

Robert A. Meyers, Ramtech Limited, Palm Desert, CA, USA

The Encyclopedia of Complexity and Systems Science series of topical volumes provides an authoritative source for understanding and applying the concepts of complexity theory, together with the tools and measures for analyzing complex systems in all fields of science and engineering. Many phenomena at all scales in science and engineering have the characteristics of complex systems, and can be fully understood only through the transdisciplinary perspectives, theories, and tools of self-organization, synergetics, dynamical systems, turbulence, catastrophes, instabilities, nonlinearity, stochastic processes, chaos, neural networks, cellular automata, adaptive systems, genetic algorithms, and so on. Examples of near-term problems and major unknowns that can be approached through complexity and systems science include: The structure, history and future of the universe; the biological basis of consciousness; the integration of genomics, proteomics and bioinformatics as systems biology; human longevity limits; the limits of computing; sustainability of human societies and life on earth; predictability, dynamics and extent of earthquakes, hurricanes, tsunamis, and other natural disasters; the dynamics of turbulent flows; lasers or fluids in physics, microprocessor design; macromolecular assembly in chemistry and biophysics; brain functions in cognitive neuroscience; climate change; ecosystem management; traffic management; and business cycles. All these seemingly diverse kinds of phenomena and structure formation have a number of important features and underlying structures in common. These deep structural similarities can be exploited to transfer analytical methods and understanding from one field to another. This unique work will extend the influence of complexity and system science to a much wider audience than has been possible to date.

Cesar E. Silva • Alexandre I. Danilenko
Editors

Ergodic Theory

A Volume in the Encyclopedia of
Complexity and Systems Science,
Second Edition

With 50 Figures

 Springer

Editors

Cesar E. Silva
Mathematics
Williams College
Williamstown, MA, USA

Alexandre I. Danilenko
B.Verkin Institute for Low Temperature
Physics and Engineering of the NAS of
Ukraine
Kharkiv, Ukraine

ISSN 2629-2327 ISSN 2629-2343 (electronic)
Encyclopedia of Complexity and Systems Science Series
ISBN 978-1-0716-2387-9 ISBN 978-1-0716-2388-6 (eBook)
<https://doi.org/10.1007/978-1-0716-2388-6>

© Springer Science+Business Media, LLC, part of Springer Nature 2023

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.

The registered company address is: 1 New York Plaza, New York, NY 10004, U.S.A.

Series Preface

The Encyclopedia of Complexity and System Science Series is a multivolume authoritative source for understanding and applying the basic tenets of complexity and systems theory as well as the tools and measures for analyzing complex systems in science, engineering, and many areas of social, financial, and business interactions. It is written for an audience of advanced university undergraduate and graduate students, professors, and professionals in a wide range of fields who must manage complexity on scales ranging from the atomic and molecular to the societal and global.

Complex systems are systems that comprise many interacting parts with the ability to generate a new quality of collective behavior through self-organization, e.g., the spontaneous formation of temporal, spatial, or functional structures. They are therefore adaptive as they evolve and may contain self-driving feedback loops. Thus, complex systems are much more than a sum of their parts. Complex systems are often characterized as having extreme sensitivity to initial conditions as well as emergent behavior that are not readily predictable or even completely deterministic. The conclusion is that a reductionist (bottom-up) approach is often an incomplete description of a phenomenon. This recognition that the collective behavior of the whole system cannot be simply inferred from the understanding of the behavior of the individual components has led to many new concepts and sophisticated mathematical and modeling tools for application to many scientific, engineering, and societal issues that can be adequately described only in terms of complexity and complex systems.

Examples of Grand Scientific Challenges which can be approached through complexity and systems science include: the structure, history, and future of the universe; the biological basis of consciousness; the true complexity of the genetic makeup and molecular functioning of humans (genetics and epigenetics) and other life forms; human longevity limits; unification of the laws of physics; the dynamics and extent of climate change and the effects of climate change; extending the boundaries of and understanding the theoretical limits of computing; sustainability of life on the earth; workings of the interior of the earth; predictability, dynamics, and extent of earthquakes, tsunamis, and other natural disasters; dynamics of turbulent flows and the motion of granular materials; the structure of atoms as expressed in the Standard Model and the formulation of the Standard Model and gravity into a Unified Theory; the structure of water; control of global infectious diseases; and also evolution and

quantification of (ultimately) human cooperative behavior in politics, economics, business systems, and social interactions. In fact, most of these issues have identified nonlinearities and are beginning to be addressed with nonlinear techniques, e.g., human longevity limits, the Standard Model, climate change, earthquake prediction, workings of the earth's interior, natural disaster prediction, etc.

The individual complex systems mathematical and modeling tools and scientific and engineering applications that comprised the *Encyclopedia of Complexity and Systems Science* are being completely updated and the majority will be published as individual books edited by experts in each field who are eminent university faculty members.

The topics are as follows:

- Agent Based Modeling and Simulation
- Applications of Physics and Mathematics to Social Science
- Cellular Automata, Mathematical Basis of
- Chaos and Complexity in Astrophysics
- Climate Modeling, Global Warming, and Weather Prediction
- Complex Networks and Graph Theory
- Complexity and Nonlinearity in Autonomous Robotics
- Complexity in Computational Chemistry
- Complexity in Earthquakes, Tsunamis, and Volcanoes, and Forecasting and Early Warning of Their Hazards
- Computational and Theoretical Nanoscience
- Control and Dynamical Systems
- Data Mining and Knowledge Discovery
- Ecological Complexity
- Ergodic Theory
- Finance and Econometrics
- Fractals and Multifractals
- Game Theory
- Granular Computing
- Intelligent Systems
- Nonlinear Ordinary Differential Equations and Dynamical Systems
- Nonlinear Partial Differential Equations
- Percolation
- Perturbation Theory
- Probability and Statistics in Complex Systems
- Quantum Information Science
- Social Network Analysis
- Soft Computing
- Solitons
- Statistical and Nonlinear Physics
- Synergetics
- System Dynamics
- Systems Biology

Each entry in each of the Series books was selected and peer reviews organized by one of our university-based book Editors with advice and consultation provided by our eminent Board Members and the Editor-in-Chief.

This level of coordination assures that the reader can have a level of confidence in the relevance and accuracy of the information far exceeding than that generally found on the World Wide Web. Accessibility is also a priority and for this reason each entry includes a glossary of important terms and a concise definition of the subject. In addition, we are pleased that the mathematical portions of our Encyclopedia have been selected by Math Reviews for indexing in MathSciNet. Also, ACM, the world's largest educational and scientific computing society, recognized our *Computational Complexity: Theory, Techniques, and Applications* book, which contains content taken exclusively from the *Encyclopedia of Complexity and Systems Science*, with an award as one of the notable Computer Science publications. Clearly, we have achieved prominence at a level beyond our expectations, but consistent with the high quality of the content!

Palm Desert, CA, USA
July 2023

Robert A. Meyers
Editor-in-Chief

Volume Preface

Ergodic theory is a branch of mathematics that studies statistical properties of dynamical systems. By statistical properties we mean properties that appear as time average for various functions computed along the orbits of the systems. The system is modeled by a map or transformation defined on the set of states. When this transformation is invertible, it defines an action of the group of integers on the set of states. Ergodic theory focuses on groups acting on a measure space in a measure-preserving (or measure-class preserving) way. Sometimes one also considers semigroup actions, and some of entries in this volume consider group and semigroup actions. The origins of ergodic theory are in the nineteenth-century work of Boltzmann on the foundations of statistical mechanics. Roughly speaking, Boltzmann hypothesized that for large systems of interacting particles in equilibrium, the “time average” is equal to the “space average.” He called this statement the ergodic hypothesis; this strong form of the hypothesis was too restrictive and was later relaxed to what lead to the formulation of the ergodic theorems. The word “ergodic” is an amalgamation of the Greek words *ergon* (work) and *hodos* (path).

Ergodic theory lies at the intersection of many areas of mathematics, including smooth dynamics, topological dynamics, statistical mechanics, differential geometry, Lie theory, probability, harmonic analysis, number theory, combinatorics, operator algebras, and group representations. Problems, techniques, and results are related to many other areas of mathematics, and ergodic theory has applications both within mathematics and to numerous other branches of science.

This encyclopedia consists of independent surveys, written by distinguished researchers, on the fundamental directions and recent developments in ergodic theory, including the asymptotic properties of measurable dynamical systems, spectral theory, entropy, ergodic theorems, joinings, recurrence, rigidity, nonsingular systems, and systems of probabilistic origin. Some surveys are devoted to the interplay between ergodic theory and topological dynamics, smooth dynamics, and operator algebras. Some entries enlighten applications to combinatorics, number theory, Sarnak’s conjecture, translation flows on translation surfaces, symbolic dynamics, pressure and equilibrium states, fractal geometry, chaos, complexity, and classification of measurable systems.

This volume is for researchers and students interested in ergodic theory and dynamics. We thank Bryna Kra who edited the first edition. We are grateful to all the researchers who have contributed to this volume.

Kharkiv, Ukraine
Williamstown, USA
July 2023

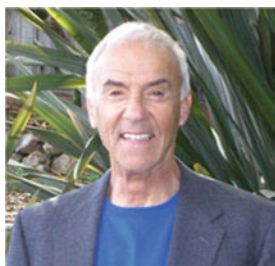
Alexandre I. Danilenko
Cesar E. Silva
Editors

Contents

Introduction to Ergodic Theory	1
Alexandre I. Danilenko and Cesar E. Silva	
Ergodic Theory: Basic Examples and Constructions	3
Matthew Nicol and Karl Petersen	
Ergodicity and Mixing Properties	35
Terrence Adams and Anthony Quas	
Ergodic Theory: Recurrence	61
Nikos Frantzikinakis and Randall McCutcheon	
Ergodic Theorems	79
Andrés del Junco	
Spectral Theory of Dynamical Systems	109
Adam Kanigowski and Mariusz Lemańczyk	
Joinings in Ergodic Theory	149
Thierry de la Rue	
Entropy in Ergodic Theory	169
Jonathan L. F. King and Kyewon Koh Park	
Isomorphism Theory in Ergodic Theory	201
Christopher Hoffman	
Dynamical Systems of Probabilistic Origin: Gaussian and Poisson Systems	217
Élise Janvresse, Emmanuel Roy, and Thierry De La Rue	
Ergodic Theory: Nonsingular Transformations	233
Alexandre I. Danilenko and Cesar E. Silva	
Sarnak's Conjecture from the Ergodic Theory Point of View	293
Joanna Kułaga-Przymus and Mariusz Lemańczyk	
Smooth Ergodic Theory	313
Amie Wilkinson	
Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces	333
Krzysztof Frączek and Corinna Ulcigrai	

Pressure and Equilibrium States in Ergodic Theory	369
Jean-René Chazottes and Gerhard Keller	
Parallels Between Topological Dynamics and Ergodic Theory ...	389
Wen Huang, Song Shao, and Xiangdong Ye	
Symbolic Dynamics	427
Brian Marcus	
Operator Ergodic Theory	459
Guy Cohen and Michael Lin	
Dynamical Systems and C^*-Algebras	491
T. Giordano and H.-C. Liao	
The Complexity and the Structure and Classification of Dynamical Systems	529
Matthew Foreman	
Ergodic Theory: Interactions with Combinatorics and Number Theory	577
Tom Ward	
Ergodic Theory on Homogeneous Spaces and Metric Number Theory	597
Dmitry Kleinbock	
Ergodic Theory: Rigidity	615
Viorel Nițică	
Chaos and Ergodic Theory	633
Jérôme Buzzi	
Ergodic Theory: Fractal Geometry	665
Jörg Schmeling	
Index	683

About the Editor-in-Chief



Robert A. Meyers

President: RAMTECH Limited

Manager, Chemical Process Technology,
TRW Inc.

Postdoctoral Fellow: California Institute of Technology

Ph.D. Chemistry, University of California at Los Angeles

B.A. Chemistry, California State University, San Diego

Biography

Dr. Meyers was manager of Energy and Environmental Projects at TRW (now Northrop Grumman) in Redondo Beach, CA, and is now president of RAMTECH Limited. He is coinventor of the Gravimelt process for desulfurization and demineralization of coal for air pollution and water pollution control and was manager of the Department of Energy project leading to the construction and successful operation of a first-of-a-kind Gravimelt Process Integrated Test Plant. Dr. Meyers is the inventor of and was project manager for the DOE-sponsored Magnetohydrodynamics Seed Regeneration Project which has resulted in the construction and successful operation of a pilot plant for production of potassium formate, a chemical utilized for plasma electricity generation and air pollution control. He also managed TRW efforts in magnetohydrodynamics electricity generating combustor and plasma channel development.

Dr. Meyers managed the pilot-scale DoE project for determining the hydrodynamics of synthetic fuels. He is a coinventor of several thermooxidative stable polymers which have

achieved commercial success as the GE PEI, Upjohn Polyimides, and Rhone-Poulenc bismaleimide resins. He has also managed projects for photochemistry, chemical lasers, flue gas scrubbing, oil shale analysis and refining, petroleum analysis and refining, global change measurement from space satellites, analysis and mitigation (carbon dioxide and ozone), hydrometallurgical refining, soil and hazardous waste remediation, novel polymers synthesis, modeling of the economics of space transportation systems, space rigidizable structures, and chemiluminescence-based devices.

He is a senior member of the American Institute of Chemical Engineers, member of the American Physical Society, and member of the American Chemical Society and has served on the UCLA Chemistry Department Advisory Board. He was a member of the joint USA-Russia working group on air pollution control and the EPA-sponsored Waste Reduction Institute for Scientists and Engineers.

Dr. Meyers has more than 20 patents and 50 technical papers in the fields of photochemistry, pollution control, inorganic reactions, organic reactions, luminescence phenomena, and polymers. He has published in primary literature journals including *Science* and the *Journal of the American Chemical Society*, and is listed in Who's Who in America and Who's Who in the World.

Dr. Meyers' scientific achievements have been reviewed in feature articles in the popular press in publications such as *The New York Times* Science Supplement and *The Wall Street Journal* as well as more specialized publications such as *Chemical Engineering* and *Coal Age*. A public service film was produced by the Environmental Protection Agency on Dr. Meyers' chemical desulfurization invention for air pollution control.

Dr. Meyers is the author or editor-in-chief of a wide range of technical books including the *Handbook of Chemical Production Processes*; the *Handbook of Synfuels Technology*; *Handbook of Petroleum Refining Processes*, now in fourth edition; the *Handbook of Petrochemical Production Processes* (McGraw-Hill), now in a second edition; the *Handbook of Energy Technology and Economics*, published by John Wiley & Sons;

Coal Structure, published by Academic Press; and *Coal Desulfurization* as well as the *Coal Handbook* published by Marcel Dekker. He served as chairman of the advisory board for *A Guide to Nuclear Power Technology: A Resource for Decision Making*, published by John Wiley & Sons, which won the Association of American Publishers Award as the best book in technology and engineering. He also served as editor-in-chief of three editions of the *Elsevier Encyclopedia of Physical Science and Technology*. Most recently, Dr. Meyers serves as editor-in-chief of the *Encyclopedia of Analytical Chemistry* as well as *Reviews in Cell Biology and Molecular Medicine* and a book series of the same name both published by John Wiley & Sons. In addition, Dr. Meyers currently serves as editor-in-chief of two Springer Nature book series, *Encyclopedia of Complexity and Systems Science* and *Encyclopedia of Sustainability Science and Technology*.

About the Volume Editors



Cesar E. Silva is the Hagey Family Professor of Mathematics at Williams College, where he has taught since 1984, and was chair of the department in 2008–2009 and 2010–2012. He received his PhD from the University of Rochester under the supervision of Dorothy Maharam, and his BS from the Pontificia Universidad Católica del Perú. He has held visiting positions at a number of universities including Maryland and Toronto. In 2015, Silva was elected a fellow of the American Mathematical Society. Silva's research interests are in ergodic theory and dynamical systems, in particular the dynamics of nonsingular and infinite measure-preserving transformations, rank-one transformations, and p-adic dynamics. Silva is the author of *Invitation to Ergodic Theory* and *Invitation to Real Analysis*, both published by the American Mathematical Society, and is co-editor of two volumes of conference proceedings, published in the AMS Contemporary Mathematics series. He was associate editor of *AMS Notices* and is the author or co-author of over 50 research articles in the general area of ergodic theory.



Dr. Alexandre I. Danilenko is currently a leading research fellow at B. Verkin Institute for Low Temperature Physics and Engineering of the National Academy of Sciences of Ukraine. He is an expert in ergodic theory and dynamical systems. Dr. Danilenko conducts research in spectral theory, orbit theory, entropy theory, theory of joinings, nonsingular dynamics, topological dynamics, etc. During his career, he has authored or co-authored more than 60 research papers published in internationally recognized mathematical journals. Alexandre I. Danilenko received

his C.Sc. degree in Mathematics at Kharkiv State University in 1991 under the supervision of V. Ya. Golodets. More than 10 years he was employed as assistant professor and associate professor at V. N. Karazin Kharkiv National University (Ukraine). He habilitated at Nicolaus Copernicus University in Torun (Poland) in 2003. Since 2002 he has been working at B. Verkin Institute for Low Temperature Physics and Engineering of the National Academy of Sciences of Ukraine.

Contributors

Terrence Adams Department of Mathematics and Statistics, State University of New York, Albany, NY, USA

Jérôme Buzzi C.N.R.S. and Université Paris-Sud, Orsay, France

Jean-René Chazottes Centre de Physique Théorique, CNRS/IP Paris, Palaiseau, France

Guy Cohen School of Electrical Engineering, Ben-Gurion University, Beer-Sheva, Israel

Alexandre I. Danilenko B.Verkin Institute for Low Temperature Physics and Engineering of the NAS of Ukraine, Kharkiv, Ukraine

Thierry De La Rue Laboratoire de Mathématiques Raphaël Salem, CNRS – Université de Rouen Normandie, Saint Étienne du Rouvray, France

Andrés del Junco Department of Mathematics, University of Toronto, Toronto, ON, Canada

Matthew Foreman University of California at Irvine, Irvine, CA, USA

Krzysztof Frączek Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland

Nikos Frantzikinakis Department of Mathematics, University of Crete, Heraklion, Greece

T. Giordano University of Ottawa, Ottawa, ON, Canada

Christopher Hoffman Department of Mathematics, University of Washington, Seattle, WA, USA

Wen Huang CAS Wu Wen-Tsun Key Laboratory of Mathematics, and Department of Mathematics, University of Science and Technology of China, Hefei, Anhui, China

Élise Janvresse Laboratoire Amiénois de Mathématique Fondamentale et Appliquée, CNRS-UMR 7352, Université de Picardie Jules Verne, Amiens, France

Adam Kanigowski Department of Mathematics, University of Maryland at College Park, College Park, MD, USA

Gerhard Keller Department Mathematik, Universität Erlangen-Nürnberg, Erlangen, Germany

Jonathan L. F. King University of Florida, Gainesville, FL, USA

Dmitry Kleinbock Department of Mathematics, Brandeis University, Waltham, MA, USA

Joanna Kułaga-Przymus Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland

Mariusz Lemańczyk Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland

H.-C. Liao University of Ottawa, Ottawa, ON, Canada

Michael Lin Department of Mathematics, Ben-Gurion University, Beer-Sheva, Israel

Brian Marcus Department of Mathematics, University of British Columbia, Vancouver, BC, Canada

Randall McCutcheon Department of Mathematics, University of Memphis, Memphis, TN, USA

Viorel Nițică West Chester University, West Chester, PA, USA
Institute of Mathematics, Bucharest, Romania

Matthew Nicol Department of Mathematics, University of Houston, Houston, TX, USA

Kyewon Koh Park Korea Institute for Advanced Study, Seoul, South Korea

Karl Petersen Department of Mathematics, University of North Carolina, Chapel Hill, NC, USA

Anthony Quas Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada

Emmanuel Roy Laboratoire Analyse, Géométrie et Applications, Université Paris 13 Institut Galilée, Villetaneuse, France

Jörg Schmeling Center for Mathematical Sciences, Lund University, Lund, Sweden

Song Shao CAS Wu Wen-Tsun Key Laboratory of Mathematics, and Department of Mathematics, University of Science and Technology of China, Hefei, Anhui, China

Cesar E. Silva Department of Mathematics, Williams College, Williamstown, MA, USA

Corinna Ulcigrai Institut für Mathematik, Universität Zürich, Zürich, Switzerland

Tom Ward School of Mathematics, University of Leeds, Leeds, UK

Amie Wilkinson Northwestern University, Evanston, IL, USA

Xiangdong Ye CAS Wu Wen-Tsun Key Laboratory of Mathematics, and
Department of Mathematics, University of Science and Technology of
China, Hefei, Anhui, China



Introduction to Ergodic Theory

Alexandre I. Danilenko¹ and Cesar E. Silva²

¹B. Verkin Institute for Low Temperature Physics and Engineering of the NAS of Ukraine, Kharkiv, Ukraine

²Department of Mathematics, Williams College, Williamstown, MA, USA

Ergodic theory traces its origins to questions in statistical mechanics about understanding motion in dynamical systems, with the motion of the planets around the sun being one of the earliest examples of a dynamical system. Since that time ergodic theory has grown into a central area of mathematics and has interacted and contributed to many areas of mathematics including harmonic analysis, number theory, probability, operator algebras, geometry, and topology. Ergodic theory is characterized by its use of measure-theoretic and probabilistic tools to study dynamical systems. At the same time, topological, differentiable, and operator tools arise in a natural way and this volume has articles that address topological and differentiable dynamics and their connections with ergodic theory. For a more detailed description of ergodic theory and its role in the modern mathematics, we refer to “Introduction to Ergodic Theory,” by Bryna Kra, from the first edition of this collection.

We owe a large debt to Bryna Kra, who was the editor of the first edition and organized the topics and the original articles. This edition would not have existed if it were not for her original vision. We have built on articles from the first edition, many of which have been updated: ▶ “[Entropy in Ergodic Theory](#)”; ▶ “[Ergodic Theory: Basic Examples and Constructions](#)”; ▶ “[Ergodic Theory: Interactions with Combinatorics and Number Theory](#)”; ▶ “[Ergodic Theory: Recurrence](#)”; ▶ “[Ergodic Theory on Homogeneous Spaces and Metric Number Theory](#)”; ▶ “[Ergodic Theory:](#)

[Nonsingular Transformations](#)”; ▶ “[Ergodicity and Mixing Properties](#)”; ▶ “[Joinings in Ergodic Theory](#)”; ▶ “[Pressure and Equilibrium States in Ergodic Theory](#)”; ▶ “[Spectral Theory of Dynamical Systems](#)”; and ▶ “[Symbolic Dynamics](#).” Some articles from the first edition are published in their original form: ▶ “[Chaos and Ergodic Theory](#)”; ▶ “[Ergodic Theory: Fractal Geometry](#)”; ▶ “[Isomorphism Theory in Ergodic Theory](#)”; and ▶ “[Smooth Ergodic Theory](#).” We are sad to report that two authors of the original edition, Andrés del Junco and Viorel Nitica, have passed away; their entries ▶ “[Ergodic Theorems](#)” and ▶ “[Ergodic Theory: Rigidity](#)” are published as in the original edition.

We are also including new surveys in this edition. The entry ▶ “[Sarnak’s Conjecture from the Ergodic Theory Point of View](#)” is devoted to impressive recent progress in studying the celebrated Sarnak conjecture (formulated in 2010) on Möbius orthogonality, which states that each topological system of zero entropy is Möbius orthogonal. This problem became a bridge between analytic number theory and dynamics. The survey covers the ergodic-theoretic aspects of this area.

Ergodic theory and topological dynamics are sibling branches of the theory of dynamical systems. The terminology used in both branches is very similar. Moreover, there are many analogous theorems. However, despite the similarities between the two theories, the proofs for “analogous” results are quite different and are not directly deduced from one another. This is the subject of the entry ▶ “[Parallels Between Topological Dynamics and Ergodic Theory](#).”

Many interesting and fundamental families of examples in ergodic theory are of probabilistic nature. The systems arising from Gaussian processes and the systems constructed from Poisson point processes are well studied in ergodic theory. These two families are of different nature, and each one is treated with specific tools. They, however, share surprisingly many

common features. This is discussed in the entry ► [“Dynamical Systems of Probabilistic Origin: Gaussian and Poisson Systems.”](#)

A classification of ergodic transformations via certain systems of parameters, such as spectrum and entropy, is the central classic problem of ergodic theory. ► [“The Complexity and the Structure and Classification of Dynamical Systems,”](#) based mostly on examples from ergodic theory, explains the methods for studying the complexity of structure, classification, and anti-classification of the systems via descriptive set theory.

Ergodic, spectral, and joining properties of translation flows on translation surfaces (and their Poincaré maps, interval exchange transformations) and smooth area-preserving locally Hamiltonian flows are discussed in ► [“Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces.”](#)

The entry ► [“Operator Ergodic Theory”](#) focuses on the asymptotic behavior of powers of a power-bounded operator in a Banach space. The asymptotic behavior is studied in different operator topologies and in various modes of convergence. That is closely related to the classic topic of ergodic theorems (see ► [“Ergodic Theorems”](#)).

Some aspects of the interplay between dynamical systems and operator algebras are discussed in ► [“Dynamical Systems and \$C^*\$ -Algebras.”](#)

While this volume contains articles that address recent areas of ergodic theory, the subject continues to develop at a fast rate, and due to space and time constraints, we have not been able to cover every possible new development.

We want to thank all the authors who have contributed to this volume, as well as the staff at Springer for their help in completing this edition.

Ergodic Theory: Basic Examples and Constructions

Matthew Nicol¹ and Karl Petersen²

¹Department of Mathematics, University of Houston, Houston, TX, USA

²Department of Mathematics, University of North Carolina, Chapel Hill, NC, USA

Article Outline

Glossary

Definition of the Subject and its Importance

Introduction

Examples

Constructions

Future Directions

Bibliography

Glossary

- A transformation T of a measure space $(X, \mathcal{B}, \mu)$ is *measure-preserving* if $\mu(T^{-1}A) = \mu(A)$ for all measurable $A \in \mathcal{B}$.
- A measure-preserving transformation $(X, \mathcal{B}, \mu, T)$ is *ergodic* if $T^{-1}(A) = A \pmod{\mu}$ implies $\mu(A) = 0$ or $\mu(A^c) = 0$ for each measurable set $A \in \mathcal{B}$.
- A measure-preserving transformation $(X, \mathcal{B}, \mu, T)$ of a probability space is *weak-mixing* if $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-i}A \cap B) - \mu(A)\mu(B)| = 0$ for all measurable sets $A, B \in \mathcal{B}$.
- A measure-preserving transformation $(X, \mathcal{B}, \mu, T)$ of a probability space is *strong-mixing* if $\lim_{n \rightarrow \infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B)$ for all measurable sets $A, B \in \mathcal{B}$.
- A continuous transformation T of a compact metric space X is *uniquely ergodic* if there is only one T -invariant Borel probability measure on X .
- Suppose $(X, \mathcal{B}, \mu)$ is a probability space. A *finite partition* $\mathcal{P}$ of X is a finite collection

of disjoint $\pmod{\mu}$, i.e., up to sets of measure 0) measurable sets $\{P_1, \dots, P_n\}$ such that $X = \bigcup P_i \pmod{\mu}$. The *entropy* of $\mathcal{P}$ with respect to μ is $H(\mathcal{P}) = -\sum_i \mu(P_i) \ln \mu(P_i)$ (other bases are sometimes used for the logarithm).

- The *metric (or measure-theoretic) entropy* of T with respect to $\mathcal{P}$ is $h_\mu(T, \mathcal{P}) = \lim_{n \rightarrow \infty} \frac{1}{n} H(\mathcal{P} \vee \dots \vee T^{-n+1}(\mathcal{P}))$, where $\mathcal{P} \vee \dots \vee T^{-n+1}(\mathcal{P})$ is the partition of X into sets of points with the same coding with respect to $\mathcal{P}$ under T^i , $i = 0, \dots, n-1$, that is, x, y are in the same set of the partition $\mathcal{P} \vee \dots \vee T^{-n+1}(\mathcal{P})$ if and only if $T^i(x)$ and $T^i(y)$ lie in the same set of the partition $\mathcal{P}$ for $i = 0, \dots, n-1$.
- The *metric entropy* $h_\mu(T)$ of $(X, \mathcal{B}, \mu, T)$ is the supremum of $h_\mu(T, \mathcal{P})$ over all finite measurable partitions $\mathcal{P}$.
- If T is a continuous transformation of a compact metric space X , then the *topological entropy* of T is the supremum of the metric entropies $h_\mu(T)$, where the supremum is taken over all T -invariant Borel probability measures.
- A system $(X, \mathcal{B}, \mu, T)$ is *loosely Bernoulli* if it is isomorphic to the first-return system to a subset of positive measure of an irrational rotation or a (positive or infinite entropy) Bernoulli system.
- Two systems are *spectrally isomorphic* if the unitary operators that they induce on their L^2 spaces are unitarily equivalent.
- A *smooth dynamical system* consists of a differentiable manifold M and a differentiable map $f: M \rightarrow M$. The degree of differentiability may be specified.
- Two submanifolds S_1, S_2 of a manifold M intersect transversely at $p \in M$ if $T_p(S_1) + T_p(S_2) = T_p(M)$.
- An (ϵ) -small C^r perturbation of a C^r map f of a manifold M is a map g such that $d_{C^r}(f, g) < \epsilon$, i.e., the distance between f and g , is less than ϵ in the C^r topology.
- A map T of an interval $I = [a, b]$ is *piecewise smooth* (C^k for $k \geq 1$) if there is a finite set of

points $a = x_1 < x_2 < \dots < x_n = b$ such that $T|(x_i, x_{i+1})$ is C^k for each i . The degree of differentiability may be specified.

- A measure μ on a measure space $(X, \mathcal{B})$ is *absolutely continuous* with respect to a measure ν on $(X, \mathcal{B})$ if $\nu(A) = 0$ implies $\mu(A) = 0$ for all measurable $A \in \mathcal{B}$.
- A Borel measure μ on a Riemannian manifold M is *absolutely continuous* if it is absolutely continuous with respect to the Riemannian volume on M .
- A measure μ on a measure space $(X, \mathcal{B})$ is equivalent to a measure ν on $(X, \mathcal{B})$ if μ is absolutely continuous with respect to ν and ν is absolutely continuous with respect to μ .

Definition of the Subject and its Importance

Measure-preserving systems are a common model of processes which evolve in time and for which the rules governing the time evolution do not change. For example, in Newtonian mechanics the planets in a solar system undergo motion according to Newton's laws of motion: The planets move, but the underlying rule governing the planets' motion remains constant. The model adopted here is to consider the time-evolution as a transformation (either a map in discrete time or a flow in continuous time) on a probability space or more generally a measure space. This is the setting of the subject called ergodic theory. Applications of this point of view include the areas of statistical physics, classical mechanics, number theory, population dynamics, statistics, information theory, and economics. The purpose of this chapter is to present a flavor of the diverse range of examples of measure-preserving transformations which have played a role in the development and application of ergodic theory and smooth dynamical systems theory. We also present common constructions involving measure-preserving systems. Such constructions may be considered a way of putting "building-block" dynamical systems together to construct examples or decomposing a complicated system into simple "building-blocks" to understand it better.

Introduction

In this chapter, we collect a brief list of some important examples of measure-preserving dynamical systems, which we denote typically by $(X, \mathcal{B}, \mu, T)$ or $(T, X, \mathcal{B}, \mu)$ or slight variations. These examples have played a formative role in the development of dynamical systems theory, either because they occur often in applications in one guise or another or because they have been useful simple models to understand certain features of dynamical systems. There is a fundamental difference in the dynamical properties of those systems which display hyperbolicity: Roughly speaking, there is some exponential divergence of nearby orbits under iteration of the transformation. In differentiable systems, this is associated with the derivative of the transformation possessing eigenvalues of modulus greater than one on a "dynamically significant" subset of phase space. Hyperbolicity leads to complex dynamical behavior such as positive topological entropy, exponential divergence of nearby orbits ("sensitivity to initial conditions") often coexisting with a dense set of periodic orbits. If ϕ, ψ are sufficiently regular functions on the phase space X of a hyperbolic measure-preserving transformation (T, X, μ) , then typically we have fast decay of correlations in the sense that

$$\left| \int_X \phi(T^n x) \psi(x) d\mu - \int \phi d\mu \int \psi d\mu \right| \leq Ca(n)$$

where $a(n) \rightarrow 0$. If $a(n) \rightarrow 0$ at an exponential rate, we say that the system has exponential decay of correlations. A theme in dynamical systems is that the time series formed by sufficiently regular observations on systems with some degree of hyperbolicity often behave statistically like independent identically distributed random variables.

At this point, it is appropriate to point out two pervasive differences between the usual probabilistic setting of a stationary stochastic process $\{X_n\}$ and the (smooth) dynamical systems setting of a time series of observations on a measure-preserving system $\{\phi \circ T^n\}$. The most crucial is that for deterministic dynamical systems the time

series is usually not an independent process, which is a common assumption in the strictly probabilistic setting. Even if some weak-mixing is assumed in the probabilistic setting, it is usually a mixing condition on the σ -algebras $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ generated by successive random variables, a condition which is not natural (and usually very difficult to check) for dynamical systems. Mixing conditions on dynamical systems are given more naturally in terms of the mixing of the sets of the σ -algebra $\mathcal{B}$ of the probability space $(X, \mathcal{B}, \mu)$ under the action of T and not by mixing properties of the σ -algebras generated by the random variables $\{\phi \circ T^n\}$. The other difference is that in the probabilistic setting, although $\{X_n\}$ satisfy moment conditions, usually no regularity properties, such as the Hölder property or smoothness, are assumed. In contrast in dynamical systems theory, the transformation T is often a smooth or piecewise smooth transformation of a Riemannian manifold X and the observation $\phi : X \rightarrow \mathbb{R}$ is often assumed continuous or Hölder. The regularity of the observation ϕ turns out to play a crucial role in proving properties such as rates of decay of correlation, central limit theorems, and so on.

An example of a hyperbolic transformation is an expanding map of the unit interval $T(x) = (2x)$ (where (x) is x modulo the integers). Here the derivative has modulus 2 at all points in phase space. This map preserves Lebesgue measure, has positive topological entropy, Lebesgue at almost every point x has a dense orbit, and periodic points for the map are dense in $[0, 1)$.

Nonhyperbolic systems are of course also an important class of examples, and in contrast to hyperbolic systems they tend to model systems of “low complexity,” for example, systems displaying quasiperiodic behavior. The simplest nontrivial example is perhaps an irrational rotation of the unit interval $[0, 1]$ given by a map $T(x) = (x + \alpha)$, $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. T preserves Lebesgue measure, and every point has a dense orbit (there are no periodic orbits), yet the topological entropy is zero and nearby points stay the same distance from each other under iteration under T .

There is a natural notion of equivalence for measure-preserving systems. We say that

measure-preserving systems $(T, X, \mathcal{B}, \mu)$ and $(S, Y, \mathcal{C}, \nu)$ are *isomorphic* if (possibly after deleting sets of measure 0 from X and Y) there is a one-to-one onto measurable map $\phi : X \rightarrow Y$ with measurable inverse ϕ^{-1} such that $\phi \circ T = S \circ \phi$ μ a.e. and $\mu(\phi^{-1}(A)) = \nu(A)$ for all $A \in \mathcal{C}$. If X, Y are compact topological spaces, we say that T is *topologically conjugate* to S if there exists a homeomorphism $\phi : X \rightarrow Y$ such that $\phi \circ T = S \circ \phi$. In this case, we call ϕ a *conjugacy*. If ϕ is C^r for some $r \geq 1$, we will call ϕ a *C^r -conjugacy* and similarly for other degrees of regularity.

We will consider $X = [0, 1] \pmod{1}$ as a representation of the unit circle $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ (under the map $x \rightarrow e^{2\pi i x}$) and similarly represent the k -dimensional torus $T^k = S^1 \times \dots \times S^1$ (k -times). If the σ -algebra is clear from the context, we will write (T, X, μ) instead of $(T, X, \mathcal{B}, \mu)$ when denoting a measure-preserving system.

Examples

Rigid Rotation of a Compact Group

If G is a compact group equipped with Haar measure and $a \in G$, then the transformation $T(x) = ax$ preserves Haar measure and is called a *rigid rotation* of G . If G is abelian and the transformation is ergodic (in this setting, transitivity implies ergodicity), then the transformation is uniquely ergodic. Such systems always have zero topological entropy.

The simplest example of such a system is a circle rotation. Take $X = [0, 1] \pmod{1}$, with

$$T(x) = (x + \alpha) \text{ where } \alpha \in \mathbb{R}.$$

Then T preserves Lebesgue (Haar) measure and is ergodic (in fact uniquely ergodic) if and only if α is irrational. Similarly, the map

$$T(x_1, \dots, x_k) = (x_1 + \alpha_1, \dots, x_k + \alpha_k),$$

where $\alpha_1, \dots, \alpha_k \in \mathbb{R}$,

preserves k -dimensional Lebesgue (Haar) measure and is ergodic (uniquely ergodic) if and

only if there are no integers $m_1, \dots, m_k$, not all 0, which satisfy $m_1 \alpha_1 + \dots + m_k \alpha_k \in \mathbb{Z}$.

Adding Machines

Let $\{k_i\}_{i \in \mathbb{N}}$ be a sequence of integers with $k_i \geq 2$. Equip each cyclic group $\mathbb{Z}_{k_i}$ with the discrete topology and form the product space $\Sigma = \prod_{i=1}^{\infty} \mathbb{Z}_{k_i}$ equipped with the product topology. An *adding machine* corresponding to the sequence $\{k_i\}_{i \in \mathbb{N}}$ is the topological space $\Sigma = \prod_{i=1}^{\infty} \mathbb{Z}_{k_i}$ together with the map

$$\sigma : \Sigma \rightarrow \Sigma$$

defined by.

$\sigma(k_1 - 1, k_2 - 1, \dots) = (0, 0, \dots)$ if each entry in the $\mathbb{Z}_{k_i}$ component is $k_i - 1$, while.
 $\sigma(k_1 - 1, k_2 - 1, \dots, k_n - 1, x_1, x_2, \dots) = (0, 0, \dots, (n \text{ times}) \dots, 0, x_1 + 1, x_2, x_3, \dots)$
 when $x_1 \neq k_{n+1} - 1$.

The map σ may be thought of as “add one and carry” and also as mapping each point to its successor in a certain order. See section “[Adic Transformations](#)” for generalizations. If each $k_i = 2$, then the system is called the *dyadic (or von Neumann-Kakutani) adding machine* or *2-odometer*. Adding machines give examples of naturally occurring minimal systems of low-orbit complexity in the sense that the topological entropy of an adding machine is zero. In fact, if f is a continuous map of an interval with zero topological entropy and S is a closed, topologically transitive invariant set without periodic orbits, then the restriction of f to S is topologically conjugate to the dyadic adding machine (Hasselblatt and Katok 2003a, Theorem 11.3.13).

We say a nonempty set Λ is an *attractor* for a map T if there is an open set U containing Λ such that $\Lambda = \bigcap_{n \geq 0} T^n(U)$ (other definitions are found in the literature). The dyadic adding machine is topologically conjugate to the Feigenbaum attractor at the limit point of period-doubling bifurcations (see section “[Unimodal Maps](#)”). Furthermore, attractors for continuous unimodal maps of the interval are either periodic orbits, transitive cycles of intervals, or Cantor sets on

which the dynamics is topologically conjugate to an adding machine (de Melo and van Strien 1993).

Interval Exchange Maps

A map $T : [0, 1] \rightarrow [0, 1]$ is an *interval exchange transformation* if it is defined in the following way. Suppose that π is a permutation of $\{1, \dots, n\}$ and $l_i > 0$, $i = 1, \dots, n$, is a sequence of subintervals of I (open or closed) with $\sum_i l_i = 1$. Define t_i by $l_i = t_i - t_{i-1}$ with $t_0 = 0$. Suppose also that σ is an n -vector with entries ± 1 . T is defined by sending the interval $t_{i-1} \leq x < t_i$ of length l_i to the interval

$$\sum_{\pi(j) < \pi(i)} l_{\pi(j)} \leq x < \sum_{\pi(j) > \pi(i)} l_{\pi(j)}$$

with orientation preserved if the i 'th entry of σ is $+1$ and orientation reversed if the i 'th entry of σ is -1 . Thus on each interval l_i , T has the form $T(x) = \sigma_i x + a_i$, where σ_i is ± 1 . If $\sigma_i = 1$ for each i , the transformation is called *orientation preserving*.

The transformation T has finitely many discontinuities (at the end points of each l_i), and modulo this set of discontinuities is smooth. T is also invertible (neglecting the finite set of discontinuities) and preserves Lebesgue measure. These maps have zero topological entropy and arise naturally in studies of polygonal billiards and more generally area-preserving flows. There are examples of minimal but nonergodic interval exchange maps (Keane 1977; Mañé 1987a).

Full Shifts and Shifts of Finite Type

Given a finite set (or alphabet) $A = \{0, \dots, d-1\}$, take $X = \Omega^+(A) = A^{\mathbb{N}}$ (or $X = A^{\mathbb{Z}}$) the sets of one-sided (two-sided) sequences, respectively, with entries from A . For example sequences in $A^{\mathbb{N}}$, have the form $x = x_0 x_1 \dots x_n \dots$. A *cylinder set* $C(y_{n_1}, \dots, y_{n_k}), y_{n_i} \in A$, of length k is a subset of X defined by fixing k entries, for example,

$$C(y_{n_1}, \dots, y_{n_k}) = \{x : x_{n_1} = y_{n_1}, \dots, x_{n_k} = y_{n_k}\}.$$

We define the set A^k to consist of all cylinders $C(y_1, \dots, y_k)$ determined by fixing the first

k entries, i.e., an element of A^k is specified by fixing the first k entries of a sequence $x_0 \dots x_k$ by requiring $x_i = y_i$, $i = 0, \dots, k$.

Let $p = (p_0, \dots, p_{d-1})$ be a probability vector: All $p_i \geq 0$ and $\sum_{i=0}^{d-1} p_i = 1$. For any cylinder $B = C(b_1, \dots, b_k) \in A^k$, define

$$g_k(B) = p_{b_1} \dots p_{b_k}. \quad (1)$$

It can be shown that these functions on A^k extend to a shift-invariant measure μ_p on $A^{\mathbb{N}}$ (or $A^{\mathbb{Z}}$) called product measure. (See the article on Measure-Preserving Systems.) The space $A^{\mathbb{N}}$ or $A^{\mathbb{Z}}$ may be given a metric by defining

$$d(x, y) = \begin{cases} 1 & \text{if } x_0 \neq y_0; \\ \frac{1}{2^{|n|}} & \text{if } x_n \neq y_n \text{ and } x_i = y_i \text{ for } |i| < n. \end{cases}$$

The *shift* $\sigma(x_0 x_1 \dots x_n \dots) = x_1 x_2 \dots x_n \dots$ is ergodic with respect to μ_p . The measure-preserving system $(\Omega, \mathcal{B}, \mu, \sigma)$ (with $\mathcal{B}$ the σ -algebra of Borel subsets of $\Omega(A)$, or its completion), is denoted by $\mathcal{B}(p)$ and is called the *Bernoulli shift* determined by p . This system models an infinite number of independent repetitions of an experiment with finitely many outcomes, the i 'th of which has probability p_i on each trial.

These systems are mixing of all orders (i.e., σ^n is mixing for all $n \geq 1$) and have countable Lebesgue spectrum (hence are all spectrally isomorphic). Kolmogorov and Sinai showed that two of them cannot be isomorphic unless they have the same entropy; Ornstein (Ornstein 1970) showed the converse. $\mathcal{B}(1/2, 1/2)$ is isomorphic to the Lebesgue-measure-preserving transformation $x \rightarrow 2x \bmod 1$ on $[0, 1]$; similarly, $\mathcal{B}(1/3, 1/3, 1/3)$ is isomorphic to $x \rightarrow 3x \bmod 1$. Furstenberg asked whether the only nonatomic measure invariant for both $x \rightarrow 2x \bmod 1$ and $x \rightarrow 3x \bmod 1$ on $[0, 1]$ is Lebesgue measure. Lyons (1988) showed that if one of the actions is K , then the measure must be Lebesgue, and Rudolph (Rudolph 1990a) showed the same thing under the weaker hypothesis that one of the actions has positive entropy. For further work on this question, see (Host 1995; Parry 1996).

This construction can be generalized to model one-step finite-state Markov stochastic processes as dynamical systems. Again let $A = \{0, \dots, d-1\}$, and let $p = (p_0, \dots, p_{d-1})$ be a probability vector. Let P be a $d \times d$ stochastic matrix with rows and columns indexed by A . This means that all entries of P are nonnegative, and the sum of the entries in each row is 1. We regard P as giving the transition probabilities between pairs of elements of A . Now we define for any cylinder $B = C(b_1, \dots, b_k) \in A^k$

$$\mu_{p,P}(B) = p_{b_1} P_{b_1 b_2} P_{b_2 b_3} \dots P_{b_{k-1} b_k}. \quad (2)$$

It can be shown that $\mu_{p,P}$ extends to a measure on the Borel σ -algebra of $\Omega^+(A)$, and its completion. (See the article on Measure-Preserving Systems.) The resulting stochastic process is a (one-step, finite-state) *Markov process*. If p and P also satisfy

$$pP = p, \quad (3)$$

then the Markov process is stationary. In this case, we call the (one- or two-sided) measure-preserving system the *Markov shift* determined by p and P .

Aperiodic and irreducible Markov chains (those for which a power of the transition matrix P has all entries positive) are strongly mixing and, in fact, are isomorphic to Bernoulli shifts (usually by means of a complicated measure-preserving recoding).

More generally, we say a set $\Lambda \subset A^{\mathbb{Z}}$ is a *subshift* if it is compact and invariant under σ . A subshift Λ is said to be of *finite type* (SFT) if there exists a $d \times d$ matrix $M = (a_{ij})$ such that all entries are 0 or 1 and $x \in \Lambda$ if and only if $a_{x_i x_{i+1}} = 1$ for all $i \in \mathbb{Z}$. Shifts of finite type are also called *topological Markov chains*. There are many invariant measures for a nontrivial shift of finite type. For example, the orbit of each periodic point is the support of an invariant measure. An important role in the theory, derived from motivations of statistical mechanics, is played by *equilibrium measures* (or *equilibrium states*) for continuous functions $\phi : \Lambda \rightarrow \mathbb{R}$, i.e., those measures μ which maximize $\{h_\sigma(\mu) + \int \phi d\mu\}$ over all shift-invariant probability measures, where

$h_\sigma(\mu)$ is the measure-theoretic entropy of σ with respect to μ . The study of full shifts or shifts of finite type has played a prominent role in the development of the hyperbolic theory of dynamical systems as physical systems with “chaotic” dynamics “typically” possess an invariant set with induced dynamics topologically conjugate to a shift of finite type (see the discussion by Smale in (Smale 1980, p. 147)). Dynamical systems in which there are transverse homoclinic connections are a common example (Guckenheimer and Holmes 1990, Theorem 5.3.5). Furthermore, in certain settings positive metric entropy implies the existence of shifts of finite type. One result along these lines is a theorem of Katok (Katok 1980). Let $h_{\text{top}}(f)$ denote the topological entropy of a map f and $h_\mu(f)$ denote metric entropy with respect to an invariant measure μ .

Theorem 4.1 (Katok) *Suppose $T : M \rightarrow M$ is a $C^{1+\epsilon}$ diffeomorphism of a closed manifold and μ is an invariant measure with positive metric entropy (i.e., $h_\mu(T) > 0$). Then for any $0 < \epsilon < h_\mu(T)$, there exists an invariant set Λ topologically conjugate to a transitive shift of finite type with $h_{\text{top}}(T|_\Lambda) > h_\mu(T) - \epsilon$.*

More Examples of Subshifts

We consider some further examples of systems that are given by the shift transformation on a subset of the set of (usually doubly infinite) sequences on a finite alphabet, usually $\{0, 1\}$. Associated with each subshift is its *language*, the set of all finite blocks seen in all sequences in the subshift. These languages are *extractive* (or *factorial*) (every subword of a word in the language is also in the language) and *insertive* (or *extendable*) (every word in the language extends on both sides to longer words in the language). In fact, these two properties characterize the languages (subsets of the set of finite-length words on an alphabet) associated with subshifts.

Prouhet-Thue-Morse

An interesting (and often rediscovered) element of $\{0, 1\}^{\mathbb{Z}_+}$ is produced as follows. Start with 0, and at each stage write down the opposite ($0' = 1$, $1' = 0$) or mirror image of what is available so far.

Or, repeatedly apply the *substitution* $0 \rightarrow 01$, $1 \rightarrow 10$.

0
0 1
0 1 10
0 1 10 0110
...

The n 'th entry is the sum, mod 2, of the digits in the dyadic expansion of n . Using Keane's *block multiplication* (Keane 1968) according to which if B is a block, $B \times 0 = B$, $B \times 1 = B'$, and $B \times (\omega_1 \dots \omega_n) = (B \times \omega_1) \dots (B \times \omega_n)$, we may also obtain this sequence as

$0 \times 01 \times 01 \times 01 \times \dots$

The orbit closure of this sequence is uniquely ergodic (there is a unique shift-invariant Borel probability measure, which is then necessarily ergodic). It is isomorphic to a skew product (see section “[Skew Products](#)”) over the von Neumann-Kakutani adding machine, or odometer (see section “[Adding Machines](#)”). *Generalized Morse systems*, that is, orbit closures of sequences like $0 \times 001 \times 001 \times 001 \times \dots$, are also isomorphic to skew products over compact group rotations.

Chacon System

This is the orbit closure of the sequence generated by the substitution $0 \rightarrow 0010$, $1 \rightarrow 1$. It is uniquely ergodic and is one of the first systems shown to be weakly mixing but not strongly mixing. It is *prime* (has no nontrivial factors) (del Junco 1978) and in fact has *minimal self joinings* (del Junco et al. 1980). It also has a nice description by means of cutting up the unit interval and stacking the pieces, using spacers (see section “[Cutting and Stacking](#)”). This system has singular spectrum. It is not known whether or not its Cartesian square is loosely Bernoulli.

Sturmian Systems

Take the orbit closure of the sequence $\omega_n = \chi_{[1-\alpha, 1)}(n\alpha)$, where α is irrational. This is a uniquely ergodic system that is isomorphic to rotation by α on the unit interval. These systems

have *minimal complexity* in the sense that the number of n -blocks grows as slowly as possible $(n + 1)$ (Coven and Hedlund 1973).

Toeplitz Systems

Fill in the blanks alternately with 0's and 1's. If done correctly, you get a uniquely ergodic system which is isomorphic to a rotation on a compact abelian group (Downarowicz 2005).

Sofic Systems

These are images of SFT's under continuous factor maps (finite codes, or block maps). They correspond to *regular languages* – languages whose words are recognizable by finite automata. These are the same as the languages defined by *regular expressions* – finite expressions built up from $\emptyset$ (empty set), ϵ (empty word), $+$ (union of two languages), (all concatenations of words from two languages), and $*$ (all finite concatenations of elements). They also have the characteristic property that the family of all *follower sets* of all blocks seen in the system is a finite family; similarly for *predecessor sets*. These are also generated by *phase-structure grammars* which are *linear*, in the sense that every production is either of the form $A \rightarrow Bw$ or $A \rightarrow w$, where A and B are variables and w is a string of terminals (symbols in the alphabet of the language).

[A *phase-structure grammar* consists of alphabets V of *variables* and A of *terminals*, a set of *productions*, which is a finite set of pairs of words (α, w) , usually written $\alpha \rightarrow w$, of words on $V \cup A$, and a *start symbol* S . The associated language consists of all words on the alphabet A of terminals which can be made by starting with S and applying a finite sequence of productions].

Sofic systems typically support many invariant measures (for example, they have many periodic points), but topologically transitive ones (those with a dense orbit) have a unique measure of maximal entropy. See (Lind and Marcus 1995).

Context-Free Systems

These are generated by phase-structure grammars in which all productions are of the form $A \rightarrow w$, where A is a variable and w is a string of variables and terminals.

Coded Systems

These are systems all of whose blocks are concatenations of some (finite or infinite) list of blocks. These are the same as the closures of increasing sequences of SFT's (Krieger 2000). Alternatively, they are the closures of the images under finite edge-labelings of irreducible countable-state topological Markov chains. They need not be context-free. Square-free languages are not coded and, in fact, do not contain any coded systems of positive entropy. See (Blanchard and Hansel 1986a, b, 1991).

Smooth Expanding Interval Maps

Take $X = [0, 1] \pmod{1}$, $m \in \mathbb{N}$, $m > 1$ and

$$T(x) = (mx)$$

Then T preserves Lebesgue measure μ (recall that T preserves μ if $\mu(T^{-1}A) = \mu(A)$ for all $A \in \mathcal{B}$). Furthermore, it can be shown that T is ergodic.

This simple map exemplifies many of the characteristics of systems with some degree of hyperbolicity. It is isomorphic to a Bernoulli shift. The map has positive topological entropy and exponential divergence of nearby orbits, and Hölder functions have exponential decay of correlations and satisfy the central limit theorem and other strong statistical properties (Boyersky and Góra 1997a).

If $m = 2$, the system is isomorphic to a model of tossing a fair coin, which is a common example of randomness. To see this, let $\mathcal{P} = \{P_0 = [0, 1/2], P_1 = [1/2, 1]\}$ be a partition of $[0, 1]$ into two subintervals. We code the orbit under T of any point $x \in [0, 1]$ by 0's and 1's by letting $x_k = i$ if $T^k x \in P_i$, $k = 0, 1, 2, \dots$. The map $\phi : X \rightarrow \{0, 1\}^{\mathbb{N}}$ which associates a point x to its *itinerary* in this way is a measure-preserving map from (X, μ) to $\{0, 1\}^{\mathbb{N}}$ equipped with the Bernoulli measure from $p_0 = p_1 = \frac{1}{2}$. The map ϕ satisfies $\phi \circ T = \sigma \circ \phi$, μ a.e. and is invertible a.e., hence is an isomorphism. Furthermore, reading the binary expansion of x is equivalent to following the orbit of x under T and noting which element of the partition $\mathcal{P}$ is entered at each time. Borel's theorem on normal numbers (base m) may be seen as a special case of the Birkhoff Ergodic Theorem in this setting.

Piecewise C^2 Expanding Maps

The main statistical features of the examples in section “[Smooth Expanding Interval Maps](#)” generalize to a broader class of expanding maps of the interval. For example:

Let $X = [0, 1]$ and let $\mathcal{P} = \{I_1, \dots, I_n\}$ ($n \geq 2$) be a partition of X into intervals (closed, half-open, or open) such that $I_i \cap I_j = \emptyset$ if $i \neq j$. Let I_i^o denote the interior of I_i . Suppose $T : X \rightarrow X$ satisfies:

- (a) For each $i = 1, \dots, n$, $T|_{I_i}$ has a C^2 extension to the closure $\bar{I}_i$ of I_i and $|T'(x)| \geq \alpha > 1$ for all $x \in I_i^o$.
- (b) $T(I_j) = \cup_{i \in P_j} I_i$ Lebesgue a.e. for some non-empty subset $P_j \subset \{1, \dots, n\}$.
- (c) For each I_j , there exists n_j such that $T^{n_j}(I_j) = [0, 1]$ Lebesgue a.e.

Then T has an invariant measure μ which is absolutely continuous with respect to Lebesgue measure m , and there exists $C > 0$ such that

$\frac{1}{C} \leq \frac{d\mu}{dm} \leq C$. Furthermore, T is ergodic with respect to μ and displays the same statistical properties listed above for the C^2 expanding maps (Boyarsky and Góra 1997a) (See the “Folklore Theorem” in the article on Measure-Preserving Systems).

More Interval Maps

Continued Fraction Map

This is the map $T : [0, 1] \rightarrow [0, 1]$ given by $Tx = 1/x \bmod 1$, and it corresponds to the shift $[0; a_1, a_2, \dots] \rightarrow [0; a_2, a_3, \dots]$ on the continued fraction expansions of points in the unit interval (a map on $\mathbb{N}^{\mathbb{N}}$). It preserves a unique finite measure equivalent to Lebesgue measure, the *Gauss measure* $dx/(\log 2)(1+x)$. It is Bernoulli with entropy $\pi^2/6 \log 2$ (in fact the natural partition into intervals is a weak Bernoulli generator; for definition and details, see (Phillips and Varadhan 1975)). By using the Ergodic Theorem, Khintchine and Lévy showed that

$$(a_1 \dots a_n)^{1/n} \rightarrow \prod_{k=1}^{\infty} \left[1 + \frac{1}{k^2 + 2k} \right]^{\log k / \log 2} \quad \text{a.e. as } n \rightarrow \infty,$$

$$\text{if } [0; a_1, \dots, a_n] = \frac{p_n}{q_n}, \text{ then } \frac{1}{n} \log q_n \rightarrow \frac{\pi^2}{12 \log 2} \quad \text{a.e.;}$$

$$\frac{1}{n} \log \left| x - \frac{p_n(x)}{q_n(x)} \right| \rightarrow \frac{\pi^2}{6 \log 2} \quad \text{a.e.};$$

and if m is Lebesgue measure (or any equivalent measure) and μ is Gauss measure, then for each interval I , $m(T^{-n}I) \rightarrow \mu(I)$, in fact exponentially fast, with a best constant 0.30366 ... See (Billingsley 1978a; Mayer 1991).

The Farey Map

This is the map $U : [0, 1] \rightarrow [0, 1]$ given by $Ux = x/(1-x)$ if $0 \leq x \leq 1/2$, $Ux = (1-x)/x$ if $1/2 \leq x \leq 1$. It is ergodic for the σ -finite infinite measure dx/x (Rényi/Parry). It is also ergodic for the *Minkowski measure* d , which is a measure of maximal entropy. This map corresponds to the shift on the *Farey tree* of rational numbers which

provide the *intermediate convergents* (best one-sided) as well as the continued fraction (best two-sided) rational approximations to irrational numbers. See (Lagarias 1991, 1992).

F-Expansions

Generalizing the continued fraction map, let $f : [0, 1] \rightarrow [0, 1]$ and let $\{I_n\}$ be a finite or infinite partition of $[0, 1]$ into subintervals. We study the map f by coding itineraries with respect to the partition $\{I_n\}$. For many examples, absolutely continuous (with respect to Lebesgue measure) invariant measures can be found and their

dynamical properties determined. See (Schweiger 1995a).

β -Shifts

This is the special case of f -expansions when $f(x) = \beta x \bmod 1$ for some fixed $\beta > 1$. This map of the interval is called the β -transformation. With a proper choice of partition, it is represented by the shift on a certain subshift of the set of all sequences on the alphabet $D = \{0, 1, \dots, \lfloor \beta \rfloor\}$, called the β -shift. A point x is expanded as an infinite series in negative powers of β with coefficients from this set; $d_\beta(x)_n = \lfloor \beta f^n(x) \rfloor$. (By convention, terminating expansions are replaced by eventually periodic ones.) A one-sided sequence on the alphabet D is in the β -shift if and only if all of its shifts are lexicographically less than or equal to the expansion $d_\beta(1)$ of 1 base β . A one-sided sequence on the alphabet D is the valid expansion of 1 for some β if and only if it lexicographically dominates all its shifts. These were first studied by Bissinger, Rényi, and Parry; there are good summaries by Bertrand-Mathis (1986) and Blanchard (1989).

For $\beta = \frac{1+\sqrt{5}}{2}$, $d_\beta(1) = 10101010\dots$

For $\beta = \frac{3}{2}$, $d_\beta(1) = 101000001\dots$ (not eventually periodic).

Every β -shift is coded.

The topological entropy of a β -shift is $\log \beta$. There is a unique measure of maximal entropy $\log \beta$.

A β -shift is a shift of finite type if and only if the β -expansion of 1 is finite. It is sofic if and only if the expansion of 1 is eventually periodic. If β is a Pisot-Vijayaragavhan number (algebraic integer all of whose conjugates have modulus less than 1), then the β -shift is sofic. If the β -shift is sofic, then β is a Perron number (algebraic integer of maximum modulus among its conjugates).

Theorem 4.2 Parry (1966) *Every strongly transitive (for every nonempty open set U , $\bigcup_{n>0} T^n U = X$) piecewise monotonic map on $[0, 1]$ is topologically conjugate to a β -transformation.*

Gaussian Systems

Consider a real-valued stationary process $\{f_k : -\infty < k < \infty\}$ on a probability space $(\Omega, \mathcal{F}, P)$. The process (and the associated measure-preserving system consisting of the shift and a shift-invariant measure on $\mathbb{R}^{\mathbb{Z}}$) is called *Gaussian* if for each $d \geq 1$, any d of the f_k form an $\mathbb{R}^d$ -valued Gaussian random variable on Ω : This means that with $E(f_k) = m$ for all k and

$$\begin{aligned} A_{ij} &= \int_{\Omega} (f_{k_i} - m)(f_{k_j} - m) dP \\ &= C(k_i - k_j) \text{ for } i, j = 1, \dots, d, \end{aligned}$$

for each Borel set $B \subset \mathbb{R}$,

$$\begin{aligned} P\{\omega : (f_{k_1}(\omega), \dots, f_{k_d}(\omega)) \in B\} &= \\ \frac{1}{2\pi^{d/2}\sqrt{\det A}} \int_B \exp\left[-\frac{1}{2}(x - (m, \dots, m))^t A^{-1}(x - (m, \dots, m))\right] dx_1 \dots dx_d. \end{aligned}$$

The function $C(k)$ is positive semidefinite and hence has an associated measure σ on $[0, 2\pi]$ such that

$$C(k) = \int_0^{2\pi} e^{ikt} d\sigma(t).$$

Theorem 4.3 *The Gaussian system is ergodic if and only if the “spectral measure” σ is continuous (i.e., nonatomic), in which case it is also weakly*

mixing. It is mixing if and only if $C(k) \rightarrow 0$ as $|k| \rightarrow \infty$. If σ is singular with respect to Lebesgue measure, then the entropy is 0; otherwise the entropy is infinite (de la Rue 1993).

For more details, see (Cornfeld et al. 1982a).

Hamiltonian Systems

This paragraph is from the article on Measure-Preserving Systems. Many systems that model physical situations can be studied by means of

Hamilton's equations. The state of the entire system at any time is specified by a vector $(q, p) \in \mathbb{R}^{2n}$, the *phase space*, with q listing the coordinates of the positions of all of the particles, and p listing the coordinates of their momenta. We assume there is a time-independent *Hamiltonian function* $H(q, p)$ such that the time development of the system satisfies *Hamilton's equations*:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \quad i = 1, \dots, n. \quad (4)$$

Often in applications, the Hamiltonian function is the sum of kinetic and potential energy:

$$H(q, p) = K(p) + U(q). \quad (5)$$

Solving these equations with initial state (q, p) for the system produces a flow $(q, p) \rightarrow T_t(q, p)$ in phase space which moves (q, p) to its position $T_t(q, p)$ t units of time later. According to *Liouville's formula* (Mañé 1987a, Theorem 3.2), this flow preserves Lebesgue measure on $\mathbb{R}^{2n}$. Calculating dH/dt by means of the Chain Rule

$$\frac{dH}{dt} = \sum_i \left(\frac{\partial H}{\partial p_i} \frac{dp_i}{dt} + \frac{\partial H}{\partial q_i} \frac{dq_i}{dt} \right)$$

and using Hamilton's equations shows that H is constant on orbits of the flow, and thus each set of constant energy $X(H_0) = \{(q, p) : H(q, p) = H_0\}$ is an invariant set. There is a natural invariant measure on a constant energy set $X(H_0)$ for the restricted flow, namely the measure given by rescaling the volume element dS on $X(H_0)$ by the factor $1/\|\nabla H\|$.

Billiard Systems

These form an important class of examples in ergodic theory and dynamical systems, motivated by natural questions in physics, particularly the behavior of gas models. Consider the motion of a particle inside a bounded region D in $\mathbb{R}^d$ with piecewise smooth (C^1 at least) boundaries. In the case of planar billiards, we have $d = 2$. The particle moves in a straight line with constant speed until it hits the boundary; at which point, it

undergoes a perfectly elastic collision with the angle of incidence equal to the angle of reflection and continues in a straight line until it next hits the boundary. It is usual to normalize and consider unit speed, as we do in this discussion for convenience. We take coordinates (x, v) given by the Euclidean coordinates in $x \in D$ together with a direction vector $v \in S^{d-1}$. A flow ϕ_t is defined with respect to Lebesgue almost every (x, v) by translating x a distance t defined by the direction vector v , taking account of reflections at boundaries. ϕ_t preserves a measure absolutely continuous with respect to Riemannian volume on (x, v) coordinates. The flow we have described is called a *billiard flow*. The corresponding *billiard map* is formed by taking the Poincaré map corresponding to the cross-section given by the boundary ∂D . We will describe the planar billiard map; the higher dimensional generalization is clear. The billiard map is a map $T : \partial D \rightarrow \partial D$, where ∂D is coordinatized by (s, θ) , $s \in [0, L]$, where L is the length of ∂D and $\theta \in (0, \pi)$ measures the angle that inward pointing vectors make with the tangent line to ∂D at s . Given a point (s, θ) , the angle θ defines an oriented line $l(s, \theta)$ which intersects ∂D in two points s and s' . Reflecting l in the tangent line to ∂D at the point s' gives another oriented line passing through s' with angle θ' (measured with respect to the angular coordinate system based at s'). The billiard map is the map $T(s, \theta) = (s', \theta')$. T preserves a measure $\mu = \sin \theta ds \times d\theta$. The billiard flow may be modeled as a suspension flow over the billiard map (see section “[Suspension Flows](#)”).

If the region D is a polygon in the plane (or polyhedron in $\mathbb{R}^d$), then ∂D consists of the faces of the polyhedron. The dynamical behavior of the billiard map or flow in regions with only flat (noncurved) boundaries is quite different to that of billiard flows or maps in regions D with strictly convex or strictly concave boundaries. The topological entropy of a flat polygonal billiard is zero. Research interest focuses on the existence and density of periodic or transitive orbits. It is known that if all the angles between sides are rational multiples of π , then there are periodic orbits (Boshernitzan 1992; Vorobets et al. 1992; Masur 1986) and they are dense in the phase space

(Boshernitzan et al. 1998). It is also known that a residual set of polygonal billiards are topologically transitive and ergodic (Zemljakov and Katok 1975; Kerckhoff et al. 1986).

On the other hand, billiard maps in which ∂D has strictly convex components are physical examples of nonuniformly hyperbolic systems (with singularities). The meaning of concave or convex varies in the literature. We will consider a billiard flow inside a circle to be a system with a strictly concave boundary, while a billiard flow on the torus from which a circle has been excised to be a billiard flow with strictly convex boundary.

The class of billiards with some strictly convex boundary components, sometimes called *dispersing billiards* or *Sinai billiards*, was introduced by Sinai (Sinai 1970) who proved many of their fundamental properties. Lazutkin (Lazutkin 1973) proved that planar billiards with generic strictly concave boundary are not ergodic. Nevertheless, Bunimovich (Bunimovič 1974; Bunimovich 1979) produced a large billiard system, *Bunimovich billiards*, with strictly concave boundary segments (perhaps with some flat boundaries as well) which were ergodic and non-uniformly hyperbolic. For more details, see (Chernov and Markarian 2006a; Katok et al. 1986; Liverani and Wojtkowski 1995; Tabachnikov 2005). We will discuss possibly the simplest example of a dispersing billiard, namely a toral billiard with a single convex obstacle. Take the torus T^2 and consider a single strictly convex subdomain S with C^∞ boundary. The domain of the billiard map is $[0, L] \times (0, \pi)$, where L is the length of ∂S . The measure $\sin(\theta)ds \times d\theta$ is preserved. If the curvature of ∂S is everywhere non-zero, then the billiard map T has positive topological entropy, periodic points are dense, and in fact the system is isomorphic to a Bernoulli shift (Gallavotti and Ornstein 1974).

KAM-Systems and Stably Nonergodic Behavior

A celebrated theorem of Kolmogorov, Arnold, and Moser (the KAM theorem) implies that the set of ergodic area-preserving diffeomorphisms of a compact surface without boundary is not dense in the C^r topology for $r \geq 4$. This has important implications, in that there are natural systems in which ergodicity is not generic. The constraint of

perturbing in the class of area-preserving diffeomorphisms is an appropriate imposition in many physical models. We will take the version of the KAM theorem as given in (Mañé 1987a, Theorem 5.1) (original references include (Kolmogorov 1954; Arnol'd 1963; Moser 1962)). An elliptic fixed point for an area-preserving diffeomorphism T of a surface M is called a *nondegenerate elliptic fixed point* if there is a local C^r , $r \geq 4$, change of coordinates h so that in polar coordinates

$$hTh^{-1}(r, \theta) = (r, \theta + \alpha_0 + \alpha_1 r) + F,$$

where all derivatives of F up to order 3 vanish, $\alpha_1 \neq 0$, and $\alpha_0 \neq 0, \frac{\pm\pi}{2}, \pi, \frac{\pm 2\pi}{3}$. A map of the form

$$\tau(r, \theta) = (r, \theta + \alpha_0 + \alpha_1 r),$$

where $\alpha_1 \neq 0$, is called a *twist map*. Note that a twist map leaves invariant the circle $r = k$, any constant k , and rotates each invariant curve by a rigid rotation $\alpha_1 r$, the magnitude of the rotation depending upon r . With respect to two-dimensional Lebesgue measure, a twist map is certainly not ergodic.

Theorem Suppose T is a volume-preserving diffeomorphism of class C^r , $r \geq 4$, of a surface M . If x is a nondegenerate elliptic fixed point, then for every $\epsilon > 0$ there exists a neighborhood U_ϵ of x and a set $U_{0,\epsilon} \subset U_\epsilon$ with the properties:

- (a) $U_{0,\epsilon}$ is a union of T -invariant simple closed curves of class C^{r-1} containing x in their interior.
- (b) The restriction of T to each such invariant curve is topologically conjugate to an irrational rotation.
- (c) $m(U_\epsilon - U_{0,\epsilon}) \leq \epsilon U_\epsilon$ where m is Lebesgue measure on M .

As a corollary, we have.

Corollary Let M be a compact surface without boundary and $\text{Diff}^r(M)$ the space of C^r area-preserving diffeomorphisms with the C^r topology. Then the set of $T \in \text{Diff}^r(M)$ which are ergodic with respect to the probability measure

determined by normalized area is not dense in $\text{Diff}^r(M)$ for $r \geq 4$.

Smooth Uniformly Hyperbolic Diffeomorphisms and Flows

Time series of measurements on deterministic dynamical systems sometimes display limit laws exhibited by independent identically distributed random variables, such as the central limit theorem, and also various mixing properties. The models of hyperbolicity we discuss in this section have played a key role in showing how this phenomenon of “chaotic behavior” arises in deterministic dynamical systems. Hyperbolic sets and their associated dynamics have also been pivotal in studies of structural stability. A smooth system is C^r *structurally stable* if a small perturbation in the C^r topology gives rise to a system which is topologically conjugate to the original. When modeling a physical system, it is desirable that slight changes in the modeling parameters do not greatly affect the qualitative or quantitative behavior of the ensemble of orbits considered as a whole. The orbit of a point may change drastically under perturbation (especially if the system has sensitive dependence on initial conditions) but the collection of all orbits should ideally be “similar” to the original unperturbed system. In the latter case, one would hope that statistical properties also vary only slightly under perturbation. Structural stability is one, quite strong, notion of stability. The conclusion of a body of work on structural stability is that a system is C^1

structurally stable if and only if it is uniformly hyperbolic and satisfies a technical assumption called strong transversality (see below for details).

Suppose M is a C^1 compact Riemannian manifold equipped with metric d and tangent space TM with norm $\|\cdot\|$. Suppose also that $U \subset M$ is a nonempty open subset and $T: U \rightarrow T(U)$ is a C^1 diffeomorphism. A compact T -invariant set $\Lambda \subset U$ is called a *hyperbolic set* if there is a splitting of the tangent space $T_p M$ at each point $p \in \Lambda$ into two invariant subspaces, $T_p M = E^u(p) \oplus E^s(p)$, and a number $0 < \lambda < 1$ such that for $n \geq 0$

$$\begin{aligned} \|D_p T^n v\| &\leq C \lambda^n \|v\| \text{ for } v \in E^s(p), \\ \|D_p T^{-n} v\| &< C \lambda^n \|v\| \text{ for } v \in E^u(p). \end{aligned}$$

The subspace E^u is called the *unstable* or *expanding subspace* and the subspace E^s the *stable* or *contracting subspace*. The stable and unstable subspaces may be integrated to produce *stable* and *unstable manifolds*

$$\begin{aligned} W^s(p) &= \{y : d(T^n p, T^n y) \rightarrow 0\} \text{ as } n \rightarrow \infty, \\ W^u(p) &= \{y : d(T^{-n} p, T^{-n} y) \rightarrow 0\} \text{ as } n \rightarrow \infty. \end{aligned}$$

The stable and unstable manifolds are immersions of Euclidean spaces of the same dimension as $E^s(p)$ and $E^u(p)$, respectively, and are of the same differentiability as T . Moreover, $T_p(W^s(p)) = E^s(p)$ and $T_p(W^u(p)) = E^u(p)$. It is also useful to define *local stable manifolds* and *local unstable manifolds* by

$$\begin{aligned} W_\epsilon^s(p) &= \{y \in W^s(p) : d(T^n p, T^n y) < \epsilon\} \text{ for all } n \geq 0, \\ W_\epsilon^u(p) &= \{y \in W^u(p) : d(T^{-n} p, T^{-n} y) < \epsilon\} \text{ for all } n \geq 0. \end{aligned}$$

Finally, we discuss the notion of strong transversality. We say a point x is *nonwandering* if for each open neighborhood U of x there exists an $n > 0$ such that $T^n(U) \cap U \neq \emptyset$. The *NW* set of nonwandering points is called the *nonwandering set*. We say a dynamical system has the *strong transversal property* if $W^s(x)$ intersects $W^u(y)$ transversely for each pair of points $x, y \in NW$.

In the C^r , $r \geq 1$ topology, Robbin (1971), de Melo (1973), and Robinson (1975, 1976) proved that dynamical systems with the strong transversal property are structurally stable, and Robinson (1973) in addition showed that strong transversality was also necessary. Mañé (1988) showed that a C^1 structurally stable diffeomorphism must be uniformly hyperbolic, and Hayashi (1997)

extended this to flows. Thus a C^1 diffeomorphism or flow on a compact manifold is structurally stable if and only if it is uniformly hyperbolic and satisfies the strong transversality condition.

Geodesic Flow on Manifold of Negative Curvature

The study of the geodesic flow on manifolds of negative sectional curvature by Hedlund and Hopf was pivotal to the development of the ergodic theory of hyperbolic systems. Suppose that M is a geodesically complete Riemannian manifold. Let $\gamma_{p,v}(t)$ be the geodesic with $\gamma_{p,v}(0) = p$ and $\dot{\gamma}_{p,v}(0) = v$, where $\dot{\gamma}_{p,v}$ denotes the derivative with respect to time t . The geodesic flow is a flow ϕ_t on the tangent bundle TM of M , $\phi_t: \mathbb{R} \times TM \rightarrow TM$, defined by

$$\phi_t(p, v) = (\gamma_{p,v}(t), \dot{\gamma}_{p,v}(t)).$$

where $(p, v) \in TM$. Since geodesics have constant speed, if $\|v\| = 1$ then $\|\dot{\gamma}_{p,v}(t)\| = 1$ for all t , and thus the unit tangent bundle $T^1M = \{(p, v) \in TM : \|v\| = 1\}$ is preserved under the geodesic flow. The geodesic flow and its restriction to the unit tangent bundle both preserve a volume form, Liouville measure. In 1934, Hedlund (1934) proved that the geodesic flow on the unit tangent bundle of a surface of strictly negative constant sectional curvature is ergodic, and in 1939 Hopf (1939) extended this result to manifolds of arbitrary dimension and strictly negative (not necessarily constant) curvature. Hopf's technique of proof of ergodicity (Hopf argument) was extremely influential and used the foliation of the tangent space into stable and unstable manifolds. For a clear exposition of this technique, and the property of absolute continuity of the foliations into stable and unstable manifolds, see (Liverani and Wojtkowski 1995). The geodesic flow on manifolds of constant negative sectional curvature is an Anosov flow (see section “Anosov Systems”). We remark that for surfaces sectional curvature is the same as Gaussian curvature. Recently, the time-one map of the geodesic flow on the unit tangent bundle of a surface with constant negative curvature, which is a partially hyperbolic system (see section “Partially

Hyperbolic Dynamical Systems”), was shown to be stably ergodic (Grayson et al. 1994), so the geodesic flow is still playing a major role in the development of ergodic theory.

Horocycle Flow

All surfaces endowed with a Riemannian metric of constant negative curvature are quotients of the upper half-plane $\mathcal{H}^+ := \{x + iy \in \mathbb{C} : y > 0\}$ with the metric $ds^2 = \frac{dx^2 + dy^2}{y^2}$, whose sectional curvature is -1 . The orientation-preserving isometries of this metric are exactly the linear fractional (also known as Möbius) transformations:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R}), \quad \left[z \in \mathcal{H}^+ \mapsto \frac{az + b}{cz + d} \in \mathcal{H}^+ \right].$$

Since each matrix $\pm I$ corresponds to the identity transformation, we consider matrices in $\mathrm{PSL}(2, \mathbb{R}) := \mathrm{SL}(2, \mathbb{R})/\{\pm I\}$.

The unit tangent bundle, $S\mathcal{H}^+$, of the upper half-plane can be identified with $\mathrm{PSL}(2, \mathbb{R})$. Then the geodesic flow corresponds to the transformations

$$t \in \mathbb{R} \mapsto \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$$

seen as acting on $\mathrm{PSL}(2, \mathbb{R})$. The unstable foliation of an element $A \in \mathrm{PSL}(2, \mathbb{R}) \cong S\mathcal{H}^+$ is given by

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} A, \quad t \in \mathbb{R},$$

and the flow along this foliation, given by

$$t \in \mathbb{R} \mapsto \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix},$$

is called the *horocycle flow*, similarly for the flow induced on the unit tangent bundle of each quotient of the upper half-plane by a discrete group of linear fractional transformations.

The geodesic and horocycle flows acting on a (finite-volume) surface of constant negative

curvature form the fundamental example of a transverse pair of actions. The geodesic flow often has many periodic orbits and many invariant measures, has positive entropy, and is in fact Bernoulli with respect to the natural measure (Ornstein and Weiss 1973), while the horocycle flow is often uniquely ergodic (Furstenberg 1973; Marcus 1975) and of entropy zero, although mixing of all orders (Marcus 1978). See (Hasselblatt and Katok 2003a) for more details.

Markov Partitions and Coding

If $(X, T, \mathcal{B}, \mu)$ is a dynamical system, then a finite partition of X always induces a coding of the orbits and a semiconjugacy with a subshift on a symbol space (it may not of course be a full conjugacy). For hyperbolic systems, a special class of partitions, *Markov partitions*, induce a conjugacy for the invariant dynamics to a subshift of finite type. A Markov partition $\mathcal{P}$ for an invariant subset Λ of a diffeomorphism T of a compact manifold M is a finite collection of sets R_i , $1 \leq i \leq n$ called *rectangles*. The rectangles have the property, for some $\epsilon > 0$, if $x, y \in R_i$ then $W_\epsilon^s(x) \cap W_\epsilon^u(y) \in R_i$. This is sometimes described as being closed under *local product structure*. We let $W^u(x, R_i)$ denote $W_\epsilon^u(x) \cap R_i$ and $W^s(x, R_i)$ denote $W_\epsilon^s(x) \cap R_i$. Furthermore, we require for all i, j :

1. Each R_i is the closure of its interior.
2. $\Lambda \subset \cup_i R_i$.
3. $R_i \cap R_j = \partial R_i \cap \partial R_j$ if $i \neq j$.
4. If $x \in R_i^o$ and $T(x) \in R_j^o$, then $W^u(T(x), R_j) \subset T(W^u(x, R_i))$ and $W^s(x, R_i) \subset T^{-1}(W^s(T(x), R_j))$.

Anosov Systems

An *Anosov diffeomorphism* (Anosov 1967) is a uniformly hyperbolic system in which the entire manifold is a hyperbolic set. Thus, an Anosov diffeomorphism is a C^1 diffeomorphism T of M with a DT -invariant splitting (which is a continuous splitting) of the tangent space $TM(x)$ at each point p into a disjoint sum

$$T_p M = E^u(p) \oplus E^s(p)$$

and $0 < \lambda < 1$, constant C such that $\|DT^n v\| < C\lambda^n \|v\|$ for all $v \in E^s(p)$ and $\|DT^{-n} w\| \leq C\lambda^n \|w\|$ for all $w \in E^u(p)$.

A similar definition holds for Anosov flows $\phi : \mathbb{R} \times M \rightarrow M$. A flow is *Anosov* if there is a splitting of the tangent bundle into flow-invariant subspaces E^u , E^s , and E^c so $D_p \phi_t E_p^s = E_{\phi_t(p)}^s$, $D_p \phi_t E_p^u = E_{\phi_t(p)}^u$, and $D_p \phi_t E_p^c = E_{\phi_t(p)}^c$, and at each point $p \in M$

$$T_p M = E_p^s \oplus E_p^u \oplus E_p^c$$

$$\|(D_p \phi_t) v\| < C\lambda^t \|v\| \text{ for } v \in E^s(p)$$

$$\|(D_p \phi_{-t}) v\| < C\lambda^t \|v\| \text{ for } v \in E^u(p)$$

for some $0 < \lambda < 1$. The tangent to the flow direction $E^c(p)$ is a neutral direction:

$$\|(D_p \phi_t) v\| = \|v\| \text{ for } v \in E^c(p).$$

Anosov proved that Anosov flows and diffeomorphisms which preserve a volume form are ergodic (Anosov 1967) and are also structurally stable. Sinai (1968) constructed Markov partitions for Anosov diffeomorphisms and hence coded trajectories via a subshift of finite type. Using ideas from statistical physics in (Sinai 1972), Sinai constructed Gibbs measures for Anosov systems. An SRB measure (see section “[Physically Relevant Measures and Strange Attractors](#)”) is a type of Gibbs measure corresponding to the potential $-\log |\det(DT|_{E^u})|$ and is characterized by the property of absolutely continuous conditional measures on unstable manifolds.

The simplest examples of Anosov diffeomorphisms are perhaps the two-dimensional hyperbolic toral automorphisms (the $n > 2$ generalization is clear). Suppose A is a 2×2 matrix with integer entries

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

such that $\det(A) = 1$ and A has no eigenvalues of modulus 1. Then A defines a transformation of the two-dimensional torus $T^2 = S^1 \times S^1$ such that if $v \in T^2$,

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix},$$

then

$$Av = \begin{pmatrix} (av_1 + bv_2) \\ (cv_1 + dv_2) \end{pmatrix}.$$

A preserves Lebesgue (or Haar) measure and is ergodic. A prominent example of such a matrix is

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix},$$

which is sometimes called the Arnold Cat Map. Each point with rational coordinates $(p_1/q_1, p_2/q_2)$ is periodic. There are two eigenvalues $\frac{1}{\lambda} < 1 < \lambda = \frac{3+\sqrt{5}}{2}$ with orthogonal eigenvectors, and the projections of the eigenspaces from $\mathbb{R}^2$ to T^2 are the stable and unstable subspaces.

Axiom A Systems

In the case of Anosov diffeomorphisms, the splitting into contracting and expanding bundles holds on the entire phase space M . A system $T: M \rightarrow M$ is an *Axiom A system* if the nonwandering set NW is a hyperbolic set and periodic points are dense in the nonwandering set. $NW \subset M$ may have Lebesgue measure zero. A set $\Lambda \subset M$ is *locally maximal* if there exists an open set U such that $\Lambda = \bigcap_{n \in \mathbb{Z}} T^n(U)$. The solenoid and horseshoe discussed below are examples of locally maximal sets. Bowen (1970) constructed Markov partitions for Axiom A diffeomorphisms. Ruelle imported ideas from statistical physics, in particular the idea of an equilibrium state and the variational principle, to the study of Axiom A systems (see Ruelle 1976, 1978). This work extended the notion of Gibbs measure and other ideas from statistical mechanics, introduced by Sinai for Anosov systems (Sinai 1972), into Axiom A systems.

One achievement of the Axiom A program was the Smale Decomposition Theorem, which breaks the dynamics of Axiom A systems into locally maximal sets and describes the dynamics on each (Bowen 1970, 1975; Smale 1980).

Theorem (Spectral Decomposition Theorem)

If T is Axiom A, then there is a unique decomposition of the nonwandering set NW of T

$$NW = \Lambda_1 \cup \dots \cup \Lambda_k$$

as a disjoint union of closed, invariant, locally maximal hyperbolic sets Λ_i such that T is transitive on each Λ_i . Furthermore, each Λ_i may be further decomposed into a disjoint union of closed sets $\Lambda_i^j, j = 1, \dots, n_i$ such that T^{n_i} is topologically mixing on each Λ_i^j and T cyclically permutes the Λ_i^j .

Horseshoe Maps This type of map was introduced by Steven Smale in the 1960s and has played a pivotal role in the development of dynamical systems theory. It is perhaps the canonical example of an Axiom A system (Smale 1980) and is conjugate to a full shift on 2 symbols. Let S be a unit square in $\mathbb{R}^2$, and let T be a diffeomorphism of S onto its image such that $S \cap T(S)$ consists of two disjoint horizontal strips S_0 and S_1 . Think of stretching S uniformly in the horizontal direction and contracting uniformly in the vertical direction to form a long thin rectangle, and then bending the rectangle into the shape of a horseshoe and laying the straight legs of the horseshoe back on the unit square S . This transformation may be realized by a diffeomorphism, and we may also require that T restricted to $T^{-1}S_i, i = 0, 1$, acts as a linear map. The restriction of T to the maximal invariant set $H = \bigcap_{i=-\infty}^{\infty} T^i S$ is a Smale horseshoe map. H is a Cantor set, the product of a Cantor set in the horizontal direction, and a Cantor set in the vertical direction. The conjugacy with the shift on two symbols is realized by mapping $x \in H$ to its itinerary with respect to the sets S_0 and S_1 under powers of T (positive and negative powers).

Solenoids The solenoid is defined on the solid torus X in $\mathbb{R}^3$ which we coordinatize as a circle of two-dimensional solid disks, so that

$$X = \{(\theta, z) : \theta \in [0, 1] \text{ and } |z| \leq 1, z \in \mathbb{C}\}$$

The transformation $T: X \rightarrow X$ is given by

$$T(\theta, z) = \left(2\theta \pmod{1}, \frac{1}{4}z + \frac{1}{2}e^{2\pi i\theta}\right)$$

Geometrically, the transformation stretches the torus to twice its length, shrinks its diameter by a factor of 4, then twists it, and doubles it over, placing the resultant object without self-intersection back inside the original solid torus. $T(X)$ intersects each disk $D_c = \{(\theta, z) : \theta = c\}$ in two smaller disks of $\frac{1}{4}$ the diameter. The transformation T contracts volume by a factor of 2 upon each application, yet there is expansion in the θ direction ($\theta \rightarrow 2\theta$). The solenoid $A = \bigcap_{n \geq 0} T^n(X)$ has zero Lebesgue measure, is T -invariant, and is (locally) topologically a line segment crossed with a two-dimensional Cantor set (A intersects each disk D_c in a Cantor set). The set A is an *attractor*, in that all points inside X limit under iteration by T upon A . $T : A \rightarrow A$ is an Axiom A system.

Partially Hyperbolic Dynamical Systems

Partially hyperbolic dynamical systems are a generalization of uniformly hyperbolic systems in that an invariant central direction is allowed but the contraction in the central direction is strictly weaker than the contraction in the contracting direction and the expansion in the central direction is weaker than the expansion in the expanding direction. More precisely, suppose M is a C^1 compact (adapted) Riemannian manifold equipped with metric d and tangent space TM with norm $\|\cdot\|$. A C^1 diffeomorphism T of M is a partially hyperbolic diffeomorphism if there is a nontrivial continuous DT -invariant splitting of the tangent space $T_p M$ at each point p into a disjoint sum

$$T_p M = E^u(p) \oplus E^c(p) \oplus E^s(p)$$

and continuous positive functions $m, M, \tilde{\gamma}, \gamma$ such that

- E^s is contracted: If $v^s \in E^s(x) \setminus \{0\}$, then $\frac{\|D_p T^n v^s\|}{\|v^s\|} \leq m(p) < 1$.
- E^u is expanded: If $v^u \in E^u(x) \setminus \{0\}$, then $\frac{\|D_p T^n v^u\|}{\|v^u\|} \geq M(p) > 1$.

- E^c is uniformly dominated by E^u and E^s : If $v^c \in E^c(x) \setminus \{0\}$, then there are numbers $\tilde{\gamma}(p), \gamma(p)$ such that $m(p) < \tilde{\gamma}(p) \leq \frac{\|D_p T^n v^c\|}{\|v^c\|} \leq \gamma(p) < M(p)$.

The notion of partial hyperbolicity was introduced by Brin and Pesin (1974) who proved existence and properties, including absolute continuity, of invariant foliations in this setting. There has been intense recent interest in partially hyperbolic systems primarily because significant progress has been made in establishing that certain volume-preserving partially hyperbolic systems are “stably ergodic,” that is, they are ergodic and under small (C^r topology) volume-preserving perturbations remain ergodic. This phenomenon had hitherto been restricted to uniformly hyperbolic systems. For recent developments, and precise statements, on stable ergodicity of partially hyperbolic systems see (Burns et al. 2001; Pugh and Shub 2004), (?Burns-Wilkinson).

Compact Group Extensions of Uniformly Hyperbolic Systems

A natural example of a partially hyperbolic system is given by a compact group extension of an Anosov diffeomorphism. If the following terms are not familiar, see section “[Constructions](#)” on standard constructions. Suppose that (T, M, μ) is an Anosov diffeomorphism, G is a compact connected Lie group, and $h : M \rightarrow G$ is a differentiable map. The skew product $F : M \times G \rightarrow M \times G$ given by

$$F(x, g) = (Tx, h(x)g)$$

has a central direction in its tangent space corresponding to the Lie algebra LG of G (as a group element h acts isometrically on G , there is no expansion or contraction) and uniformly expanding and contracting bundles corresponding to those of the tangent space of $T : M \rightarrow M$. Thus, $T(M \times G) = E^u \oplus LG \oplus E^s$.

Time-One Maps of Anosov Flows

Another natural context in which partial hyperbolicity arises is in time-one maps of uniformly hyperbolic flows. Suppose $\phi_t : \mathbb{R} \times M \rightarrow M$ is an Anosov flow. The diffeomorphism $\phi_1 : M \rightarrow M$ is

a partially hyperbolic diffeomorphism with central direction given by the flow direction. There is no expansion or contraction in the central direction.

Nonuniformly Hyperbolic Systems

The assumption of uniform hyperbolicity is quite restrictive, and few “chaotic systems” found in applications are likely to exhibit uniform hyperbolicity. A natural weakening of this assumption, and one that is nontrivial and greatly extends the applicability of the theory, is to require the hyperbolic splitting (no longer uniform) to hold only at almost every point of phase space. A systematic theory was built by Pesin (1976, 1977) on the assumption that the system has nonzero Lyapunov exponents μ almost everywhere, where μ is Lebesgue-equivalent invariant probability measure. Recall that a number λ is a *Lyapunov exponent* for $p \in M$ if $\|D_p T^n v\| \sim e^{\lambda n}$ for some unit vector $v \in T_p M$. Oseledec’s theorem (Oseledec 1968) (see also Walters 1982a, p. 232), which is also called the Multiplicative Ergodic Theorem, implies that if T is a C^1 diffeomorphism of M , then for any T -invariant ergodic measure μ almost every point has well-defined Lyapunov exponents. One of the highlights of Pesin theory is the following structure theorem: If $T: M \rightarrow M$ is a $C^{1+\epsilon}$ diffeomorphism with a T -invariant Lebesgue-equivalent Borel measure μ such that T has nonzero Lyapunov exponents with respect to μ , then T has at most a countable number of ergodic components $\{C_i\}$ on each of which the restriction of T is either Bernoulli or Bernoulli times a rotation (by which we mean the support of $\mu_i = \mu|_{C_i}$ consists of a finite number n_i of sets $\{S_1^i, \dots, S_{n_i}^i\}$ cyclically permuted and T^{n_i} is Bernoulli when restricted to each S_j^i) (Pesin 1977; Young 1993). This structure theorem has been generalized to SRB measures with nonzero Lyapunov exponents (Ledrappier 1984; Pesin 1977).

Physically Relevant Measures and Strange Attractors

(This paragraph is from the article on Measure-Preserving Systems.) For Hamiltonian systems and other volume-preserving systems, it is natural

to consider ergodicity (and other statistical properties) of the system with respect to Lebesgue measure. In dissipative systems, a measure equivalent to Lebesgue may not be invariant (for example, the solenoid). Nevertheless, Lebesgue measure has a distinguished role since sampling by experimenters is done with respect to Lebesgue measure. The idea of a physically relevant measure μ is that it determines the statistical behavior of a positive Lebesgue measure set of orbits, even though the support of μ may have zero Lebesgue measure. An example of such a situation in the uniformly hyperbolic setting is the solenoid Λ , where the attracting set Λ has Lebesgue measure zero and is (locally) topologically the product of a two-dimensional Cantor set and a line segment. Nevertheless, Λ determines the behavior of all points in a solid torus in $\mathbb{R}^3$. More generally, suppose that $T: M \rightarrow M$ is a diffeomorphism on a compact Riemannian manifold and that m is a version of Lebesgue measure on M , given by a smooth volume form. Although Lebesgue measure m is a distinguished physically relevant measure, m may not be invariant under T , and the system may even be volume contracting in the sense that $m(T^n A) \rightarrow 0$ for all measurable sets A . Nevertheless, an experimenter might observe long-term “chaotic” behavior whenever the state of the system gets close to some compact invariant set X which attracts a positive m -measure of orbits in the sense that these orbits limit on X . Possibly $m(X) = 0$, so that X is effectively invisible to the observer except through its effects on orbits not contained in X . The dynamics of T restricted to X can in fact be quite complicated – maybe a full shift, or a shift of finite type, or some other complicated topological dynamical system. Suppose there is a T -invariant measure μ supported on X such that for all continuous functions $\phi: M \rightarrow \mathbb{R}$

$$\frac{1}{n} \sum_{k=0}^{n-1} \phi \circ T^k(x) \rightarrow \int_X \phi d\mu, \quad (6)$$

for a positive m -measure of points $x \in M$. Then the long-term equilibrium dynamics of an observable set of points $x \in M$ (i.e., a set of points of positive m measure) is described by (X, T, μ) . In this situation, μ is described as a *physical*

measure. There has been a great deal of research on the properties of systems with attractors supporting physical measures.

In the dissipative nonuniformly hyperbolic setting, the theory of “physically relevant” measures is best developed in the theory of SRB (for Sinai, Ruelle, and Bowen) measures. These dynamically invariant measures may be supported on a set of Lebesgue measure zero yet determine the asymptotic behavior of points in a set of positive Lebesgue measure.

If T is a diffeomorphism of M and μ is a T -invariant Borel probability measure with positive Lyapunov exponents which may be integrated to unstable manifolds, then we call μ an *SRB measure* if the conditional measure μ induced on the unstable manifolds is absolutely continuous with respect to the Riemannian volume element on these manifolds. The reason for this definition is technical but can be gleaned from the following observation. Suppose that the diffeomorphism has no zero Lyapunov exponents with respect to μ . Since T is a diffeomorphism, this implies T has negative Lyapunov exponents as well as positive Lyapunov exponents and corresponding local stable manifolds as well as local unstable manifolds. Suppose that a T -invariant set A consists of a union of unstable manifolds and is the support of an ergodic SRB measure μ and that $\phi : M \rightarrow \mathbb{R}$ is a continuous function. Since μ has absolutely continuous conditional measures on unstable manifolds with respect to conditional Lebesgue measure on the unstable manifolds, almost every point x in the union of unstable manifolds U satisfies

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \phi \circ T^j(x) = \int \phi d\mu \quad (7)$$

If $y \in W_{\epsilon}^s(x)$ for such an $x \in U$, then $d(T^n x, T^n y) \rightarrow \infty$ and hence (Eq. 7) implies

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} \phi \circ T^j(y) = \int \phi d\mu$$

Furthermore, if the holonomy between unstable manifolds defined by sliding along stable

manifolds is absolutely continuous (takes sets of zero Lebesgue measure on W^u to sets of zero Lebesgue measure on W^u), there is a positive Lebesgue measure of points (namely an unstable manifold and the union of stable manifolds through it) satisfying (Eq. 7). Thus an SRB measure with absolutely continuous holonomy maps along stable manifolds is a physically relevant measure. If the stable foliation possesses this property, it is called *absolutely continuous*. An Axiom A attractor for a C^2 diffeomorphism is an example of an SRB attractor (Bowen 1975; Ruelle 1976, 1978; Sinai 1972). The examples we have given of SRB measures and attractors and measures have been uniformly hyperbolic.

Recently, much progress has been made in understanding the statistical properties of non-uniformly hyperbolic systems by using a tower (see section “[Induced Transformations](#)”) to construct SRB measures. We refer to Young’s original papers (Young 1998, 1999), to the book by Baladi (2000a), and to (Young 1993) for a recent survey on SRB measures in the nonuniformly setting.

Unimodal Maps

Maps of an interval to itself are simple examples of non-uniformly hyperbolic systems that have played an important role in the development of dynamical systems theory. Suppose $I \subset \mathbb{R}$ is an interval; for simplicity, we take $I = [0, 1]$. A *unimodal map* is a map $T : [0, 1] \rightarrow [0, 1]$ such that there exists a point $0 < c < 1$ and

- T is C^2 .
- $T'(x) > 0$ for $x < c$, $T'(x) < 0$ for $x > c$.
- $T'(c) = 0$.

Such a map is clearly not uniformly expanding, as $|T'(x)| < 1$ for points in a neighborhood of c . The family of maps $T_{\mu}(x) = \mu x(1-x)$, $0 < \mu \leq 4$, is a family of unimodal maps with $c = 1/2$ and $T_2(1/2) = 1/2$, $T_4(1/2) = 1$.

We could have taken the interval I to be $[-1, 1]$ or indeed any interval with an obvious modification of the definition above. A well-studied family of unimodal maps in this setting is the *logistic family* $f_a : [-1, 1] \rightarrow [-1, 1]$, $f_a(x) = 1 - ax^2$, $a \in (0, 2)$. The families are equivalent under a

smooth coordinate change, so statements about one family may be translated into statements about the other.

Unimodal maps are studied because of the insights they offer into transitions from regular or periodic to chaotic behavior as a parameter (e.g., μ or a) is varied, the existence of absolutely continuous measures, and rates of decay of correlations of regular observations for nonuniformly hyperbolic systems.

A result of Jakobson (1981) and Benedicks and Carleson (1985) implies that in the case of the logistic family there is a positive Lebesgue measure set of a such that f_a has an absolutely continuous ergodic invariant measure μ_a . It has been shown by Young (1999) and Keller and Nowicki (1992) that if f_a is mixing with respect to μ_a then the decay of correlations for Lipschitz observations on I is exponential. It is also known that the set of a such that f_a is mixing with respect to μ_a has positive Lebesgue measure. There is a well-developed theory concerning the bifurcations the maps T_μ undergo as μ varies (Collet and Eckmann 1980a). We briefly describe the period-doubling route to chaos in the family $T_\lambda(x) = \lambda x(1 - x)$. For a good account, see (Hasselblatt and Katok 2003a). We let c_λ denote the fixed point $\frac{\lambda-1}{\lambda}$. For $3 < \lambda \leq 1 + \sqrt{6}$, all points in $[0, 1]$ except for 0, c_λ , and their preimages are attracted to a unique periodic orbit $O(p_\lambda)$ of period 2. There is a monotone sequence of parameter values λ_n ($\lambda_1 = 3$) such that for $\lambda_n < \lambda \leq \lambda_{n+1}$, T_λ has a unique attracting periodic orbit $O(\lambda_n)$ of period 2^n and for each $k = 1, 2, \dots, n - 1$ a unique repelling orbit of period 2^k . All points in the interval $[0, 1]$ except for the repelling periodic orbits and their preimages are attracted to the attracting periodic orbit of period 2^n . At $\lambda = \lambda_n$, the periodic orbit $O(\lambda_n)$ undergoes a period-doubling bifurcation. Feigenbaum (1978) found that the limit $\delta = \frac{\lambda_n - \lambda_{n-1}}{\lambda_{n+1} - \lambda_n} \sim 4.699 \dots$ exists and that in a wide class of unimodal maps this period-doubling cascade occurs and the differences between successive bifurcation parameters give the same limiting ratio, an example of *universality*. At the end of the period-doubling cascade at a parameter $\lambda_\infty \sim 3.569 \dots$, T_{λ_∞} has an invariant Cantor set C (the Feigenbaum attractor) which is

topologically conjugate to the dyadic adding machine coexisting with isolated repelling orbits of period 2^n , $n = 0, 1, 2, \dots$. There is a unique repelling orbit of period 2^n for $n \geq 1$ along with two fixed points. The Cantor set is the ω -limit set for all points that are not periodic or preimages of periodic orbits. C is the set of accumulation points of periodic orbits. Despite this picture of incredible complexity, the topological entropy is zero for $\lambda \leq \lambda_\infty$. For $\lambda > \lambda_\infty$, the map T_λ has positive topological entropy and infinitely many periodic orbits whose periods are not powers of 2. For each $\lambda \geq \lambda_\infty$, T_λ possesses an invariant Cantor set which is repelling for $\lambda > \lambda_\infty$. We say that T_λ is *hyperbolic* if there is only one attracting periodic orbit and the only recurrent sets are the attracting periodic orbit, repelling periodic orbits, and possibly a repelling invariant Cantor set. It is known that the set of $\lambda \in [0, 4]$ for which T_λ is hyperbolic is open and dense (Graczyk and Świątek 1997). Remarkably, by Jakobson's result (Jakobson 1981) there is also a positive Lebesgue measure set of parameters λ for which T_λ has an absolutely continuous invariant measure μ_λ with a positive Lyapunov exponent.

Intermittent Maps

Maps of the unit interval $T: [0, 1] \rightarrow [0, 1]$ which are expanding except at the point $x = 0$, where they are locally $x \sim x + x^{1+\alpha}$, $\alpha > 0$, have been extensively studied both for the insights they give into rates of decay of correlations for non-uniformly hyperbolic systems (hyperbolicity is lost at the point $x = 0$, where the derivative is 1) and for their use as models of intermittent behavior in turbulence (Manneville and Pomeau 1980). A fixed point where the derivative is 1 is sometimes called an *indifferent fixed point*. It is a model of *intermittency* in the sense that orbits close to 1 will stay close for many iterates (since the expansion is very weak there), and hence a time series of observations will be quite uniform for long periods of time before displaying chaotic type behavior after moving away from the indifferent fixed into that part of the domain where the map is uniformly expanding.

A particularly simple model (Liverani et al. 1999) is provided by

$$T(x) = \begin{cases} x(1 + 2^\alpha x^\alpha) & \text{if } x \in [0, 1/2]; \\ 2x - 1 & \text{if } x \in [1/2, 1]. \end{cases}$$

For $\alpha = 0$, the map is uniformly expanding and Lebesgue measure is invariant. In this case, the rate of decay of correlations for Hölder observations is exponential. For $0 < \alpha < 1$, the map has an SRB measure μ_α with support the unit interval. For $\alpha \geq 1$, there are no absolutely continuous invariant probability measures though there are σ -finite absolutely continuous measures. Upper and lower polynomial bounds on the rate of decay of observations on such maps have been given as a function of $0 < \alpha < 1$ and the regularity of the observable. For details, see (Huyi 2004; Liverani et al. 1999; Sarig 2002).

Hénon Diffeomorphisms

The Hénon family of diffeomorphisms was introduced and studied as Poincaré maps for the Lorenz system of equations. It is a two-parameter two-dimensional family which shares many characteristics with the logistic family and for small $b > 0$ may be considered a two-dimensional “perturbation” of the logistic family. The parametrized mapping is defined as

$$T_{a,b}(x, y) = (1 - ax^2 + y, bx),$$

so $T_{a,b} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ with $0 < a < 2$ and $b > 0$. Benedicks and Carleson (1991) showed that for a positive-measure set of parameters (a, b) , $T_{a,b}$ has a topologically transitive attractor $\Lambda_{a,b}$. Benedicks and Young (1993) later proved that for a positive-measure set of parameters (a, b) , $T_{a,b}$ has a topologically transitive SRB attractor $\Lambda_{a,b}$ with SRB measure $\mu_{a,b}$ and that $(T_{a,b}, \Lambda_{a,b}, \mu_{a,b})$ is isomorphic to a Bernoulli shift.

Complex Dynamics

Complex dynamics is concerned with the behavior of rational maps

$$\frac{\alpha_1 z^d + \alpha_2 z^{d-1} + \dots + \alpha_{d+1}}{\beta_1 z^d + \beta_2 z^{d-1} + \dots + \beta_{d+1}}$$

of the extended complex plane $\overline{\mathbb{C}}$ to itself, in which the domain is $\mathbb{C}$ completed with the point at infinity (called the Riemann sphere). Recall that a family F of meromorphic functions is called *normal* on a domain D if every sequence possesses a subsequence that converges uniformly (in the spherical metric $\overline{\mathbb{C}} \sim S^2$) on compact subsets of D . A family is normal at a point $z \in \overline{\mathbb{C}}$ if it is normal on a neighborhood of z . The *Fatou set* $F(R) \subset \overline{\mathbb{C}}$ of a rational map $R : \overline{\mathbb{C}} \rightarrow \overline{\mathbb{C}}$ is the set of points $z \in \overline{\mathbb{C}}$ such that the family of forward iterates $\{R^n\}_{n \geq 0}$ is normal at z . The *Julia set* $J(R)$ is the complement of the Fatou set $F(R)$. The Fatou set is open, and hence the Julia set is a closed set. Another characterization in the case $d > 1$ is that $J(R)$ is the closure of the set of all repelling periodic orbits of $R : \overline{\mathbb{C}} \rightarrow \overline{\mathbb{C}}$. Both $F(R)$ and $J(R)$ are invariant under R . The dynamics of greatest interest is the restriction $R : J(R) \rightarrow J(R)$. The Julia set often has a complicated fractal structure. In the case that $Ra(z) = z^2 - a$, $a \in \mathbb{C}$, the *Mandelbrot set* is defined as the set of a for which the orbit of the origin 0 is bounded. The topology of the Mandelbrot set has been the subject of intense research. The study of complex dynamics is important because of the fascinating and complicated dynamics displayed and also because techniques and results in complex dynamics have direct implications for the behavior of one-dimensional maps. For more details, see (Carleson and Gamelin 1993a).

Infinite Ergodic Theory

We may also consider a measure-preserving transformation (T, X, μ) of a measure space such that $\mu(X) = \infty$. For example, X could be the real line equipped with Lebesgue measure. This setting also arises with compact X in applications. For example, suppose $T : [0, 1] \rightarrow [0, 1]$ is the simple model of intermittency given in section “[Intermittent Maps](#)” and $\gamma \in (1, 2)$. Then T possesses an absolutely continuous invariant measure μ with support $[0, 1]$, but $\mu([0, 1]) = \infty$. The Radon-Nikodym derivative of μ with respect to Lebesgue measure m exists but is not in $L^1(m)$.

In this setting, we say a measurable set A is a *wandering set* for T if $\{T^{-n}A\}_{n=0}^\infty$ are disjoint. Let

$D(T)$ be the measurable union of the collection of wandering sets for T . The transformation T is *conservative* with respect to μ if $(X \setminus D(T)) = X \pmod{\mu}$ (for more details, see the recent survey by A. Danilenko and C. Silva in this volume). It is usually necessary to assume T conservative with respect to μ to say anything interesting about its behavior. For example, if $T(x) = x + \alpha$, $\alpha > 0$ is a translation of the real line, then $D(T) = X$. The definition of ergodicity in this setting remains the same: T is *ergodic* if $A \in \mathcal{B}$, and $T^{-1}A = A \pmod{\mu}$ implies that $\mu(A) = 0$ or $\mu(A^c) = 0$. However, the equivalence of ergodicity of T with respect to μ and the equality of time and space averages for $L^1(\mu)$ functions no longer holds. Thus, in general μ ergodic does not imply that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \phi \circ T^i(x) = \int_X \phi d\mu \quad \mu \text{ a.e. } x \in X$$

for all $\phi \in L^1(\mu)$. In the example of the intermittent map with $\gamma \in (1, 2)$, the orbit of Lebesgue almost every $x \in X$ is dense in X , yet the fraction of time spent near the indifferent fixed point $x = 0$ tends to one for Lebesgue almost every $x \in X$. In fact, it may be shown (Aaronson 1997, Section 2.4) that when $\mu(x) = \infty$ there are no constants $a_n > 0$ such that

$$\lim_{n \rightarrow \infty} \frac{1}{a_n} \sum_{i=0}^{n-1} \phi \circ T^i(x) = \int_X \phi d\mu \quad \mu \text{ a.e. } x \in X$$

Nevertheless, it is sometimes possible to obtain distributional limits, rather than almost sure limits, of Birkhoff sums under suitable normalization. We refer the reader to Aaronson's book (Aaronson 1997) for more details.

Constructions

We give examples of some of the standard constructions in dynamical systems. Often these constructions appear in modeling situations (for example, skew products are often used to model

systems which react to inputs from other systems, and continuous time systems are often modeled as suspension flows over discrete-time dynamics) or to reduce systems to simpler components (often a factor system or induced system is simpler to study). Unless stated otherwise, in the following we discuss measure-preserving transformations on Lebesgue spaces (see the article on Measure-Preserving Systems).

Products

Given measure-preserving systems $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$, their *product* consists of their completed product measure space with the transformation $T \times S : X \times Y \rightarrow X \times Y$ defined by $(T \times S)(x, y) = (Tx, Sy)$ for all $(x, y) \in X \times Y$. Neither ergodicity nor transitivity is in general preserved by taking products, for example, the product of an irrational rotation on the unit circle with itself is not ergodic. For a list of which mixing properties are preserved under the taking of products, see (Walters 1982a). Given any countable family of measure-preserving transformations on probability spaces, their direct product is defined similarly.

Factors

We say that a measure-preserving system $(Y, \mathcal{C}, \nu, S)$ is a *factor* of a measure-preserving system $(X, \mathcal{B}, \mu, T)$ if (possibly after deleting a set of measure 0 from X) there is a measurable onto map $\phi : X \rightarrow Y$ such that

$$\begin{aligned} \phi^{-1}C &\subset \mathcal{B}, \\ \phi T &= S\phi, \text{ and} \\ \mu T^{-1} &= \nu. \end{aligned} \tag{8}$$

For Lebesgue spaces, factors of $(X, \mathcal{B}, \mu, T)$ correspond perfectly with T -invariant complete sub- σ -algebras of $\mathcal{B}$. According to Rokhlin's theory of Lebesgue spaces (Rohlin 1952) (see the article on Measure-Preserving Systems), factors also correspond perfectly to certain kinds of partitions of X . A factor map $\phi : X \rightarrow Y$ between Lebesgue spaces is an *isomorphism* if and only if it has a measurable inverse, or equivalently $\phi^{-1}C = \mathcal{B}$ up to sets of measure 0.

Skew Products

If $(X, \mathcal{B}, \mu, T)$ is a measure-preserving system, $(Y, \mathcal{C}, \nu)$ is a measure-space, and $\{S_x : x \in X\}$ is a family of measure-preserving maps $Y \rightarrow Y$ such that the map that takes (x, y) to $S_x y$ is jointly measurable in the two variables x and y ; then we may define a *skew product system* consisting of the product measure space of X and Y equipped with product measure $\mu \times \nu$ together with the measure-preserving map $T \ltimes S : X \times Y \rightarrow X \times Y$ defined by

$$(T \ltimes S)(x, y) = (Tx, S_x y). \quad (9)$$

The space Y is called the *fiber* of the skew product and the space X the *base*. Sometimes in the literature, the word skew product has a more general meaning and refers to the structure $(T \ltimes S)(x, y) = (Tx, S_x y)$ (without any assumption of measure-preservation), where the action of the map on the fiber Y is determined or “driven” by the map $T : X \rightarrow X$.

Some common examples of skew products include the following:

Random Dynamical Systems

Suppose S_x is considered a (random) choice of a mapping $Y \rightarrow Y$ from the set $\{S_x : x \in X\}$. We suppose $T : X \rightarrow X$ to be the full shift. Then the projection onto Y of the orbits of $(Tx, S_x y)$ give the orbits of a point $y \in Y$ under a random composition of maps $S_{T^n x} \circ \dots \circ S_{Tx} \circ S_x$. More generally, we could consider the choice of maps S_x that are composed to come from any ergodic dynamical system, (T, X, μ) to model the effect of perturbations by a stationary ergodic “noise” process.

Group Extensions of Dynamical Systems

Suppose Y is a group, ν is a measure on Y invariant under a left group action, and $S_x y := g(x)y$ is given by a group-valued function $g : X \rightarrow Y$. In this setting, g is often called a *cocycle*, since upon defining $g^{(n)}(x)$ by $(T \ltimes S)^{(n)}(x, y) = (T^n x, g^{(n)}(x)y)$ we have a cocycle relation, namely $g^{(m+n)}(x) = g^{(m)}(T^n x)g^{(n)}(x)$. Group extensions arise often in modeling systems with symmetry (Field and Nicol 2004). Common examples are provided by

a random composition of matrices from a group of matrices (or more generally from a set of matrices which may form a group or not).

Induced Transformations

Since by the Poincaré Recurrence Theorem a measure-preserving transformation $(T, X, \mu, \mathcal{B})$ on a probability space is recurrent, given any set B of positive measure, the return-time function

$$n_B(x) = \inf \{n \geq 1 : T^n x \in B\} \quad (10)$$

is finite μ a.e. We may define the *first-return map* by

$$T_B x = T^{n_B(x)} x. \quad (11)$$

Then (after perhaps discarding as usual a set of measure 0) $T_B : B \rightarrow B$ is a measurable transformation which preserves the probability measure $\mu_B = \mu/\mu(B)$. The system $(B, \mathcal{B} \cap B, \mu_B, T_B)$ is called an *induced*, *first-return*, or *derived transformation*. If $(T, X, \mu, \mathcal{B})$ is ergodic, then $(B, \mathcal{B} \cap B, \mu_B, T_B)$ is ergodic, but the converse is not in general true.

The construction of the transformation T_B allows us to represent the forward orbit of points in B via a *tower* or *skyscraper* over B . For each $n = 1, 2, \dots$, let

$$B_n = \{x \in B : n_B(x) = n\}. \quad (12)$$

Then $\{B_1, B_2, \dots\}$ form a partition of B , which we think of as the bottom floor or base of the tower. The next floor is made up of $TB_2, TB_3, \dots$, which form a partition of $TB \setminus B$, and so on. All these sets are disjoint. A *column* is a part of the tower of the form $B_n \cup TB_n \cup \dots \cup T^{n-1}B_n$ for some $n = 1, 2, \dots$. The action of T on the entire tower is pictured as mapping each x not at the top of its column straight up to the point Tx above it on the next level, and mapping each point on the top level to $T^{n_B} x \in B$. An equivalent way to describe the transformation on the tower is to write for each n and $j < n$, $T^j B_n$ as $\{(x, j) : x \in B_n\}$, and then the transformation F on the tower becomes

$$F(x, l) = \begin{cases} (x, l+1) & \text{if } l < n_B(x) - 1; \\ (T^{n_B(x)}x, 0) & \text{if } l = n_B(x) - 1. \end{cases}$$

If T preserves a measure μ , then F preserves $\mu \times dl$, where l is counting measure so that the measure $\mu \times dl$ can be naturally lifted to the tower.

Sometimes, the process of inducing yields an induced map which is easier to analyze (perhaps it has stronger hyperbolicity properties) than the original system. Sometimes also, it is possible to “lift” ergodic or statistical properties from an induced system to the original system, so the process of inducing plays an important role in the study of statistical properties of dynamical systems (Melbourne and Török 2004).

It is possible to generalize the tower construction and relax the condition that $n_B(x)$ is the first-return time function. We may take a measurable set $B \subset X$ of positive μ measure and define for almost every point $x \in B$ a *height or ceiling* function $R : B \rightarrow \mathbb{N}$ and take a countable partition $\{X_n\}$ of B into the sets on which R is constant. We define the *tower* as the set $\Delta := \{(x, l) : x \in B, 0 \leq l < R(x)\}$ and the *tower map* $F : \Delta \rightarrow \Delta$ by

$$F(x, l) = \begin{cases} (x, l+1) & \text{if } l < R(x) - 1; \\ (T^{R(x)}x, 0) & \text{if } l = R(x) - 1. \end{cases}$$

In this setting, if $\int_B R(x) d\mu < \infty$, we may define an F -invariant probability measure on Δ as $\frac{\mu}{C(R, B)} \times dl$, where dl is counting measure and $C(R, B)$ is the normalizing constant $C(R, B) = \mu(B) \int_B R(x) d\mu$. This viewpoint is connected with the construction of systems by cutting and stacking – see section “[Cutting and Stacking](#).”

Suspension Flows

The tower construction has an analogue in which the height function R takes values in $\mathbb{R}$ rather than $\mathbb{N}$. Such towers are commonly used to model dynamical systems with continuous time parameter. Let (T, X, μ) be a measure-preserving system and $R : X \rightarrow (0, \infty)$ a measurable “ceiling” function on X . The set

$$X^R = \{(x, t) : 0 \leq R(x) < t\}, \quad (13)$$

with measure ν given locally by the product of μ on X with Lebesgue measure m on $\mathbb{R}$, is a measure space in a natural way. If μ is a finite measure and R is integrable with respect to μ , then ν is a finite measure. We define an action of $\mathbb{R}$ on X^R by letting each point x flow at unit speed up the vertical lines $\{(x, t) : 0 \leq t < R(x)\}$ under the graph of R until it hits the ceiling, then jump to Tx , and so on. More precisely, defining $R_n(x) = R(x) + \dots + R(T^n x)$,

$$T_s(x, t) = \begin{cases} (x, s+t) & \text{if } 0 \leq s+t < R(x), \\ (Tx, s+t-f(x)) & \text{if } R(x) \leq s+t < R(x) + R(Tx) \\ \dots & \\ (T^n x, s+t - [R(x) + \dots + R(T^{n-1}x)]) & \text{if } R_{n-1}(x) \leq s+t < R_n(x). \end{cases} \quad (14)$$

Ergodicity of (T, X, μ) implies the ergodicity of (T_s, X^R, ν) .

Cutting and Stacking

Many of the most interesting examples in ergodic theory have been constructed by this method; in fact, because of Rokhlin’s Lemma (see section “[Rokhlin’s Lemma](#)”) every ergodic measure-preserving transformation on a Lebesgue space

is isomorphic to one constructed by cutting and stacking. We could mention especially the von Neumann-Kakutani adding machine (or 2-odometer) (section “[Adding Machines](#)”), the Chacon weakly mixing but not strongly mixing system (section “[Chacon System](#)”), Ornstein’s mixing rank one examples (see Nadkarni 1998a, p. 160 ff.), and many more.

We construct a Lebesgue measure-preserving transformation T on an interval X (bounded or maybe unbounded) by defining it as a translation on each of a pairwise disjoint countable collection of subintervals. The construction proceeds by stages, at each stage defining T on an additional part of X , until eventually T is defined a.e.

At each stage, X is represented as a *tower*, which is defined to be a disjoint union of *columns*. A *column* is defined to be a finite disjoint union of intervals of equal length, which are numbered from 0, for the “floor,” to the last one, for the “roof,” and which we picture as lying each above the preceding-numbered one. T is defined on each *level* of a column (i.e., each interval in the column) except the roof by mapping it by translation to the next higher interval in the column.

At stage 0, we have just one column, consisting of all of X as the floor, and T is not defined anywhere. To pass from one stage to the next, the columns are *cut* and *stacked*. This means that each column is divided, by vertical cuts, into a disjoint union of subcolumns of equal height (but maybe not equal width), and then some of these subcolumns are stacked above others (of the same width) so as to form a new tower. This allows the definition of T to be extended to some parts of X that were previously tops of towers, since they now may have levels above them (Sometimes columns of height 1 are thought of as forming a reservoir for “spacers” to be inserted between subcolumns that are being stacked). If the measure of the union of the tops of the columns tends to 0, eventually T becomes defined a.e. This description in words can be made precise with cumbersome notation, but the process can also be given a neater graphical description, which we sketch in the next section.

Adic Transformations

A.M. Vershik has introduced a family of models, called *adic* or *Bratteli-Vershik* transformations, into ergodic theory and dynamical systems. One begins with a graph which is arranged in levels, finitely many vertices on each level, with connections only from each level to the adjacent ones. The space X consists of the set of all infinite paths in this graph; it is a compact metric space in a natural way. We are given an order on the set of

edges into each vertex, and then X is partially ordered as follows: x and y are comparable if they agree from some point on, in which case we say that $x < y$ if at the last level n , where they traverse different edges, the edge x_n of x is smaller than the edge y_n of y . A map T is defined by letting Tx be the smallest y that is larger than x , if there is one. In nice situations, T is a homeomorphism after defining it and its inverse on perhaps countably many maximal and minimal elements. Invariant measures can sometimes be defined by assigning weights to edges, which are then multiplied to define the measure of each cylinder set. This is a nice combinatorial way to present the cutting and stacking method of constructing m.p.t.'s, allows for more convenient analysis of questions such as orbit equivalence, and leads to the construction of many interesting examples, such as those based on the Pascal or Euler graphs (Bailey et al. 2006; Frick and Petersen; Méla and Petersen 2005). Odometers and generalizations are natural examples of adic systems. Vershik showed that in fact every ergodic measure-preserving transformation on a Lebesgue space is isomorphic to a uniquely ergodic adic transformation. See (Vershik and Livshits 1992).

Rokhlin's Lemma

The following result is the fundamental starting point for many constructions in ergodic theory, from representing arbitrary systems in terms of cutting and stacking or adic systems, to constructing useful partitions and symbolic codings of abstract systems, to connecting convergence theorems in abstract ergodic theory with those in harmonic analysis. It allows us to picture arbitrarily long stretches of the action of a measure-preserving transformation as a translation within the set of integers. In the ergodic nonatomic case, the statement follows readily from the construction of derivative transformations.

Lemma 5.1 (Rokhlin's Lemma) *Let $T : X \rightarrow X$ be a measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$. Suppose that $(X, \mathcal{B}, \mu)$ is nonatomic and $T : X \rightarrow X$ is ergodic, or, more generally, $(T, X, \mathcal{B}, \mu)$ is aperiodic: that is to say, in the set $\{x \in X : \text{there is } n \in \mathbb{N} \text{ such that } T^n x =$*

$x\}$ of periodic points has measure 0. Then given $n \in \mathbb{N}$ and $\epsilon > 0$, there is a measurable set $B \subset X$ such that the sets $B, TB, \dots, T^{n-1}B$ are pairwise disjoint and $\mu(\cap_{k=0}^{n-1} T^k B) > 1 - \epsilon$.

Inverse Limits

Suppose that for each $i = 1, 2, \dots$ we have a Lebesgue probability space $(X_i, \mathcal{B}_i, \mu_i)$ and a measure-preserving transformation $T_i : X_i \rightarrow X_i$. Suppose also that for each $i \leq j$ there is a factor map $\phi_{ji} : (T_j, X_j, \mathcal{B}_j, \mu_j) \rightarrow (T_i, X_i, \mathcal{B}_i, \mu_i)$, such that each ϕ_{ji} is the identity on X_j and $\phi_{ji} \phi_{kj} = \phi_{ki}$ whenever $k \geq j \geq i$. Let

$$X = \left\{ x \in \prod_{i=1}^{\infty} X_i : \phi_{ji} x_j = x_i \text{ for all } j \geq i \right\}. \quad (15)$$

For each j , let $\pi_j : X \rightarrow X_j$ be the projection defined by $\pi_j x = x_j$.

Let $\mathcal{B}$ be the smallest σ -algebra of subsets of X which contains all the $\pi_j^{-1} \mathcal{B}_j$. Define μ on each $\pi_j^{-1} \mathcal{B}_j$ by

$$\mu(\pi_j^{-1} B) = \mu_j(B) \text{ for all } B \in \mathcal{B}_j. \quad (16)$$

Because $\phi_{ji} \pi_j = \pi_i$ for all $j \geq i$, the $\pi_j^{-1} \mathcal{B}_j$ are increasing, and so their union is an algebra. The set function μ can, with some difficulty, be shown to be countably additive on this algebra: Since we are dealing with Lebesgue spaces, by means of measure-theoretic isomorphisms it is possible to replace the entire situation by compact metric spaces and continuous maps, then use regularity of the measures involved – see (Parthasarathy 2005, p. 137 ff.). Thus, by Carathéodory's Theorem (see the article on Measure-Preserving Systems), μ extends to all of $\mathcal{B}$.

Define $T : X \rightarrow X$ by $T(x_j) = (T_j x_j)$. Then $(T, X, \mathcal{B}, \mu)$ is a measure-preserving system which has all the $(T_j, X_j, \mathcal{B}_j, \mu_j)$ as factors, and any system that factors onto all the $(T_j, X_j, \mathcal{B}_j, \mu_j)$ also factors onto $(T, X, \mathcal{B}, \mu)$.

Natural Extension

The natural extension is a way to produce an invertible system from a noninvertible system. The original system is a factor of its natural extension, and its orbit structure and ergodic properties are captured by the natural extension, as will be seen from its construction. Let $(T, X, \mathcal{B}, \mu)$ be a measure-preserving transformation of a Lebesgue probability space. Define

$$\Omega := \{(x_0, x_1, x_2, \dots) : x_n = T(x_{n+1}), x_n \in X, n = 0, 1, 2, \dots\}$$

with $\sigma : \Omega \rightarrow \Omega$ defined by $\sigma((x_0, x_1, x_2, \dots)) = (T(x_0), x_0, x_1, \dots)$. The map σ is invertible on Ω . Given the invariant measure μ , we define the

invariant measure for the natural extension $\tilde{\mu}$ on Ω by defining it first on cylinder sets $C(A_0, A_1, \dots, A_k)$ by

$$\tilde{\mu}(C(A_0, A_1, \dots, A_k)) = \mu(T^{-k}(A_0) \cap T^{-k+1}(A_1) \dots \cap T^{-k+i}(A_i) \cap \dots \cap A_k)$$

and then extending it to Ω using Kolmogorov's extension theorem. We think of $(x_0, x_1, x_2, \dots)$ as being an inverse branch of $x_0 \in X$ under the mapping $T : X \rightarrow X$. The maps $\sigma, \sigma^{-1} : \Omega \rightarrow \Omega$ are ergodic with respect to $\tilde{\mu}$ if $(T, X, \mathcal{B}, \mu)$ is ergodic (Walters 1982a). If $\pi : \Omega \rightarrow X$ is projection onto the first component, i.e., $\pi(x_0, \dots, x_n,$

$\dots) = x_0$, then $\pi \circ \sigma^n(x_0, \dots, x_n, \dots) = T^n(x_0)$ for all x_0 and thus the natural extension yields all information about the orbits of X under T .

The natural extension is an inverse limit. Let $(X, \mathcal{B}, \mu)$ be a Lebesgue probability space and $T : X \rightarrow X$ a map such that $T^{-1} \mathcal{B} \subset \mathcal{B}$ and $\mu T^{-1} = \mu$. For each $i = 1, 2, \dots$ let $(T_i, X_i, \mathcal{B}_i, \mu_i) = (T, X, \mathcal{B},$

μ), and $\phi_{ji} = T^{j-i}$ for each $j > i$. Then the inverse limit $(\widehat{T}, \widehat{X}, \widehat{\mathcal{B}}, \widehat{\mu})$ of this system is an *invertible* measure-preserving system which is the natural extension of $(T, X, \mathcal{B}, \mu)$. We have

$$\widehat{T}^{-1}(x_1, x_2, \dots) = (x_2, x_3, \dots). \quad (17)$$

The original system $(T, X, \mathcal{B}, \mu)$ is a factor of $(\widehat{T}, \widehat{X}, \widehat{\mathcal{B}}, \widehat{\mu})$ (using any π_i as the factor map), and any factor mapping from an invertible system onto $(T, X, \mathcal{B}, \mu)$ consists of a factor mapping onto $(\widehat{T}, \widehat{X}, \widehat{\mathcal{B}}, \widehat{\mu})$ followed by projection onto the first coordinate.

Joinings

Given measure-preserving systems $(T, X, \mathcal{B}, \mu)$ and $(S, Y, \mathcal{C}, \nu)$, a *joining* of the two systems is a $T \times S$ -invariant measure P on their product measurable space that projects to μ and ν , respectively, under the projections of $X \times Y$ to X and Y , respectively. That is, if $\pi_1 : X \times Y \rightarrow X$ is the projection onto the first component, i.e., $\pi_1(x, y) = x$, then $P(\pi_1^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}$ and similarly for $\pi_2 : X \times Y \rightarrow Y$.

This concept is the ergodic-theoretic version of the notion in probability theory of a *coupling*. The product measure $\mu \times \nu$ is always a joining of the two systems. If product measure is the only joining of the two systems, then we say that they are disjoint and write $X \perp Y$ (Furstenberg 1967). If $\mathcal{D}$ is any family of systems, we write $\mathcal{D}^\perp$ for the family of all measure-preserving systems which are disjoint from every system in $\mathcal{D}$. Extensive recent accounts of the use of joinings in ergodic theory are in (Glasner 2003a; Rudolph 1990b; Thouvenot 1995a).

Future Directions

The basic examples and constructions presented here are idealized, and many of the underlying assumptions (such as uniform hyperbolicity) are seldom satisfied in applications, yet they have given important insights into the behavior of real-world physical systems. The ergodic

properties of dynamical systems will continue to be an active research area for the foreseeable future.

The directions will include, among others, the following:

1. Establishing statistical and ergodic properties under weakened dependence assumptions. There is a “hierarchy” of probabilistic limit theorems including, among others, ergodicity (akin to the strong law of large numbers for integrable observations); central limit theorem (distributional convergence to a Gaussian for the scaled Birkhoff sums of an observable with finite second moment); law of the iterated logarithm (almost sure rate of growth of scaled Birkhoff sums); and the almost sure invariance principle (a strong form of approximation by Brownian motion). Some observables on some systems exhibit the same limit laws as iid processes, for example, Hölder observables on smooth hyperbolic systems satisfy the almost sure invariance principle. In other settings, the determinism of the system plays a key role, for example, the return time statistics of a system to a periodic orbit may best be described by a compound Poisson process rather than a Poisson law (Hirata 1993). An important strand of research is determining conditions, usually dynamical and mixing conditions, on a system to determine the form of statistics that observables on the system will display. This research has also produced a rich class of counterexamples. A good reference is (Melbourne and Török 2004).
2. The study of systems which display “anomalous statistics.” By “anomalous statistics” is usually meant non-Gaussian limit laws such as the convergence of a scaled observable to a stable law rather than a Gaussian. This arises basically in two ways in a dynamical system: (1) a nonintegrable observable on a fast mixing system; (2) a regular (usually Hölder) observable on a slowly mixing dynamical system. For example, scaled Birkhoff sums of a nonintegrable observable such as $d(x, x_0)^{-1}$ on a rapidly mixing system such as the doubling map will converge to a stable law as will scaled

- Birkhoff sums of a Hölder observable on a slowly mixing intermittent map with non-summable decay of correlations (Gouëzel 2004).
3. The study of the stability and typicality of ergodic behavior and mixing in dynamical systems. This topic is often called “stable ergodicity.” Ergodicity and other mixing and statistical properties are often stable to perturbation of the underlying parameters of a dynamical system. This strand of research has produced many surprising examples of stability. For example, skew-products of a hyperbolic dynamical system with a circle S^1 fiber are typically stable to perturbation. A good review is (Burns et al. 2001).
 4. The study of statistical recurrence, which is closely related to extremal statistics on a dynamical system. One theme of research into the statistical properties of mixing dynamical systems is that limit laws for observables on the system tend to resemble those of iid processes. However, this viewpoint breaks down when considering, for example, return times to a periodic point or the extreme value statistics of observables maximized at a periodic point. Such a situation leads to clustering of returns or extrema. This study is important for applications of dynamical systems theory to, for example, climate and population models. A recent review is (Boyarsky and Góra 1997a).
 5. Sequential and nonstationary dynamical systems. In physical systems, a common situation is where the rules governing the time evolution of the system change over time. This may be modeled by a sequential dynamical system in which the maps or flows applied to the system change over time. The term sequential dynamical system was coined by Berend and Bergelson (1984), but the idea predates this work. This strand of work is important in the study of fluid mixing. In this setting, if the sequential maps or flows are hyperbolic/mixing then “self-norming” statistical properties may occur. Although there is usually not an invariant measure, statistical properties with respect to a reference measure, such as Lebesgue measure, sometimes occur. An influential work in this area is (Conze and Raugi 2007).
 6. The ergodic theory of infinite-dimensional systems and relations to partial differential equations (Young 2017).
 7. Advances in number theory (see the sections on Szemerédi and Ramsey theory); research into models with nonsingular rather than invariant measures.
 8. Infinite-measure systems. In this setting, ergodic properties are considered with respect to a σ -finite measure rather than a probability measure. An excellent reference for this topic is (Aaronson 1997).
- Other chapters in this Encyclopedia discuss in more detail these and other topics in which future research directions will be active and fruitful.

Bibliography

Primary Literature

- Aaronson J (1997) An introduction to infinite ergodic theory, *Mathematical Surveys and Monographs*, vol. 50. American Mathematical Society, Providence, MR1450400 (99d:28025)
- Anosov DV (1967) Geodesic flows on closed Riemannian manifolds of negative curvature. *Trudy Mat Inst Steklov* 90:209 (Russian). MR0224110 (36 #7157)
- Arnol'd VI (1963) Small denominators and problems of stability of motion in classical and celestial mechanics. *UspehiMat Nauk* 18(114):91–192 (Russian). MR0170705 (30 #943)
- Bailey S, Keane M, Petersen K, Salama IA (2006) Ergodicity of the adic transformation on the Euler graph. *Math Proc Cambridge Philos Soc* 141(2):231–238. MR2265871 (2007m:37010)
- Baladi V (2000a) Positive transfer operators and decay of correlations, *Advanced Series in Nonlinear Dynamics*, vol. 16. World Scientific Publishing Co. Inc., River Edge. MR1793194 (2001k:37035)
- Benedicks M, Carleson L (1985) On iterations of $1 - ax(X, \mathcal{B}, \mu)$ on $(-1, 1)$, *Ann Math* 122:1, 1–25. MR799250 (87c:58058)
- Benedicks M, Carleson L (1991) The dynamics of the Hénon map. *Ann Math* 133(1):73–169. MR1087346 (92d:58116)
- Benedicks M, Young L-S (1993) Sinai-Bowen-Ruelle measures for certain Hénon maps. *Invent Math* 112(3):541–576. MR1218323 (94e:58074)
- Berend D, Bergelson V (1984) Ergodic and mixing sequences of transformations. *Ergodic Theory Dynam*

- Syst 4(3):353–366. <https://doi.org/10.1017/S0143385700002509>. MR776873
- Bertrand-Mathis A (1986) Développement en base θ ; répartition modulo un de la suite $(x\theta^n(X, \mathcal{B}, \mu))_{n \geq 0}$; langages codés et θ -shift. Bull Soc Math France 114(3): 271–323 (French, with English summary). MR878240 (88e:11067)
- Billingsley P (1978a) Ergodic theory and information. Robert E. Krieger Publishing Co., Huntington. Reprint of the 1965 original. MR524567 (80b:28017)
- Blanchard F (1989) β -expansions and symbolic dynamics. Theor Comput Sci 65(2):131–141. MR1020481 (90j:54039)
- Blanchard F, Hansel G (1986a) Systèmes codés. Theor Comput Sci 44(1):17–49. (French, with English summary). MR858689 (88m:68029)
- Blanchard F, Hansel G (1986b) Systèmes codés et limites de systèmes sofiques. C R Acad Sci Paris Sér I Math 303(10):475–477. (French, with English summary). MR865864 (87m:94009)
- Blanchard F, Hansel G (1991) Sofic constant-to-one extensions of subshifts of finite type. Proc Am Math Soc 112(1):259–265. MR1050016 (91m:54050)
- Boshernitzan MD (1992) Billiards and rational periodic directions in polygons. Am Math Monthly 99(6): 522–529. MR1166001 (93d:51043)
- Boshernitzan M, Galperin G, Krüger T, Troubetzkoy S (1998) Periodic billiard orbits are dense in rational polygons. Trans Am Math Soc 350(9):3523–3535. MR1458298 (98k:58179)
- Bowen R (1970) Markov partitions for Axiom A diffeomorphisms. Am J Math 92:725–747. MR0277003 (43 #2740)
- Bowen R (1975) Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Lecture Notes in Mathematics, vol 470. Springer-Verlag, Berlin. MR0442989 (56 #1364)
- Boyarsky A, Góra PL (1997a) Laws of chaos, probability and its applications. Birkhäuser Boston Inc., Boston. Invariant measures and dynamical systems in one dimension. MR1461536 (99a:58102)
- Brin MI, Pesin JB (1974) Partially hyperbolic dynamical systems. Izv Akad Nauk SSSR Ser Mat 38:170–212. (Russian). MR0343316 (49 #8058)
- Bunimovič LA (1974) The ergodic properties of certain billiards. Funkcional Anal i Priložen 8(3):73–74. (Russian). MR0357736 (50 #10204)
- Bunimovich LA (1979) On the ergodic properties of nowhere dispersing billiards. Commun Math Phys 65(3):295–312. MR530154 (80h:58037)
- Burns K, Pugh C, Shub M, Wilkinson A (2001) Recent results about stable ergodicity, Smooth ergodic theory and its applications (Seattle, WA, 1999). In: Proceedings of Symposia in Pure Mathematics, vol. 69, Am Math Soc, Providence, pp. 327–366. MR1858538 (2002m:37042)
- Carleson L, Gamelin TW (1993a) Complex dynamics, Universitext: Tracts in Mathematics. Springer-Verlag, New York. MR1230383 (94h:30033)
- Chernov N, Markarian R (2006a) Chaotic billiards, Mathematical Surveys and Monographs, vol 127. American Mathematical Society, Providence. MR2229799 (2007f:37050)
- Collet P, Eckmann J-P (1980a) Iterated maps on the interval as dynamical systems, Progress in Physics, vol 1. Birkhäuser, Boston. MR613981 (82j:58078)
- Conze J-P, Raugi A (2007) Limit theorems for sequential expanding dynamical systems on $[0, 1]$. In: Ergodic theory and related fields, Contemporary Mathematics, vol. 430, Am Math Soc, Providence, pp. 89–121, <https://doi.org/10.1090/conm/430/08253>, (to appear in print). MR2331327
- Cornfeld IP, Fomin SV, Sinai YG (1982a) Ergodic theory, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 245, Springer-Verlag, New York. Translated from the Russian by A. B. Sosinskii. MR832433 (87f:28019)
- Coven EM, Hedlund GA (1973) Sequences with minimal block growth. Math Syst Theory 7:138–153. MR0322838 (48 #1199)
- de la Rue T (1993) Entropie d'un système dynamique gaussien: cas d'une action de $Z(X, \mathcal{B}, \mu)$. C R Acad Sci Paris Sér I Math 317(2):191–194 (French, with English and French summaries). MR1231420 (94c:28022)
- de Melo W (1973) Structural stability of diffeomorphisms on two-manifolds. Invent Math 21:233–246. MR0339277 (49 #4037)
- de Melo W, van Strien S (1993) One-dimensional dynamics, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol 25. Springer-Verlag, Berlin. MR1239171 (95a:58035)
- del Junco A (1978) A simple measure-preserving transformation with trivial centralizer. Pacific J Math 79(2): 357–362. MR531323 (80i:28034)
- del Junco A, Rahe M, Swanson L (1980) Chacon's automorphism has minimal self-joinings. J Anal Math 37: 276–284. MR583640 (81j:28027)
- Downarowicz T (2005) Survey of odometers and Toeplitz flows, Algebraic and topological dynamics. In: Contemporary Mathematics, vol. 385, Am Math Soc, Providence, pp. 7–37. MR2180227 (2006f:37009)
- Feigenbaum MJ (1978) Quantitative universality for a class of nonlinear transformations. J Statist Phys 19(1):25–52. MR0501179 (58 #18601)
- Field M, Nicol M (2004) Ergodic theory of equivariant diffeomorphisms: Markov partitions and stable ergodicity. Mem Am Math Soc 169(803):viii+100. MR2045641 (2005g:37041)
- Frick SB, Petersen K Random permutations and unique fully supported ergodicity for the Euler adic transformation. Ann Inst H Poincaré Prob Stat to appear
- Furstenberg H (1967) Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. Math Syst Theory 1:1–49. MR0213508 (35 #4369)
- Furstenberg H (1973) The unique ergodicity of the horocycle flow. Recent advances in topological

- dynamics (Proc. Conf., Yale Univ., New Haven, Conn., 1972; in honor of Gustav Arnold Hedlund), Springer, Berlin, pp. 95–115. Lecture Notes in Math, vol. 318. MR0393339 (52 #14149)
- Gallavotti G, Ornstein DS (1974) Billiards and Bernoulli schemes. *Commun Math Phys* 38:83–101. MR0355003 (50 #7480)
- Glasner E (2003a) Ergodic Theory via Joinings, Mathematical Surveys and Monographs, vol 101. American Mathematical Society, Providence. MR1958753 (2004c:37011)
- Gouëzel S (2004) Central limit theorem and stable laws for intermittent maps. *Probab Theory Relat Fields* 128(1): 82–122. <https://doi.org/10.1007/s00440-003-0300-4>. MR2027296
- Graczyk J, Świątek G (1997) Generic hyperbolicity in the logistic family. *Ann Math* 146(1):1–52. MR1469316 (99b:58079)
- Guckenheimer J, Holmes P (1990) Nonlinear oscillations, dynamical systems, and bifurcations of vector fields, Applied Mathematical Sciences, vol. 42. Springer-Verlag, New York. Revised and corrected reprint of the 1983 original. MR1139515 (93e:58046)
- Grayson M, Pugh C, Shub M (1994) Stably ergodic diffeomorphisms. *Ann Math* 140(2):295–329. MR1298715 (95g:58128)
- Hasselblatt B, Katok A (2003a) A first course in dynamics. Cambridge University Press, New York. With a panorama of recent developments. MR1995704 (2004f:37001)
- Hayashi S (1997) Connecting invariant manifolds and the solution of the $C(X, \mathcal{B}, \mu)$ stability and Ω -stability conjectures for flows. *Ann Math* 145(1):81–137. MR1432037 (98b:58096)
- Hedlund GA (1934) On the metrical transitivity of the geodesics on closed surfaces of constant negative curvature. *Ann Math.* 35(4):787–808. MR1503197
- Hirata M (1993) Poisson law for Axiom A diffeomorphisms. *Ergodic Theory Dyn Syst* 13(3):533–556. <https://doi.org/10.1017/S0143385700007513>. MR1245828
- Hopf E (1939) Statistik der geodätischen Linien in Mannigfaltigkeiten negative Krümmung. *Ber Verh Sächs Akad Wiss Leipzig* 91:261–304. (German). MR0001464 (1,243a)
- Host B (1995) Nombres normaux, entropie, translations. *Israel J Math* 91(1–3):419–428. (French, with English summary). MR1348326 (96g:11092)
- Huyi H (2004) Decay of correlations for piecewise smooth maps with indifferent fixed points. *Ergodic Theory Dyn Syst* 24(2):495–524. MR2054191 (2005a:37064)
- Jakobson MV (1981) Absolutely continuous invariant measures for one-parameter families of one-dimensional maps. *Commun Math Phys* 81(1):39–88. MR630331 (83j:58070)
- Katok A (1980) Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst Hautes Études Sci Publ Math* 51:137–173. MR573822 (81i:28022)
- Katok A, Strelcyn J-M, Ledrappier F, Przytycki F (1986) Invariant manifolds, entropy and billiards; smooth maps with singularities, Lecture Notes in Mathematics, vol. 1222, Springer-Verlag, Berlin. MR872698 (88k:58075)
- Keane M (1968) Generalized Morse sequences. *Z Wahrscheinlichkeitstheorie Verw Gebiete* 10: 335–353. MR0239047 (39 #406)
- Keane M (1977) Non-ergodic interval exchange transformations. *Israel J Math* 26(2):188–196. MR0435353 (55 #8313)
- Keller G, Nowicki T (1992) Spectral theory, zeta functions and the distribution of periodic points for Collet-Eckmann maps. *Commun Math Phys* 149(1):31–69. MR1182410 (93i:58123)
- Kerckhoff S, Masur H, Smillie J (1986) Ergodicity of billiard flows and quadratic differentials. *Ann Math* 124(2):293–311. MR855297 (88f:58122)
- Kolmogorov AN (1954) On conservation of conditionally periodic motions for a small change in Hamilton's function. *Dokl Akad Nauk SSSR (NS)* 98:527–530. (Russian). MR0068687 (16,924c)
- Krieger W (2000) On subshifts and topological Markov chains, Numbers, information and complexity (Bielefeld, 1998). Kluwer Academic Publishers, Boston, pp. 453–472. MR1755380 (2001g:37010)
- Lagarias JC (1991) The Farey shift, manuscript
- Lagarias JC (1992) Number theory and dynamical systems. In: The unreasonable effectiveness of number theory (Orono, ME, 1991), Proceedings of Symposia in Pure Mathematics, vol. 46, American Mathematical Society, Providence, pp. 35–72. MR1195841 (93m:11143)
- Lazutkin VF (1973) Existence of caustics for the billiard problem in a convex domain. *Izv Akad Nauk SSSR Ser Mat* 37:186–216. (Russian). MR0328219 (48 #6561)
- Ledrappier F (1984) Propriétés ergodiques des mesures de Sinaï. *Inst Hautes Études Sci Publ Math* 59:163–188. (French). MR743818 (86f:58092)
- Lind D, Marcus B (1995) An introduction to symbolic dynamics and coding. Cambridge University Press, Cambridge. MR1369092 (97a:58050)
- Liverani C, Saussol B, Vienti S (1999) A probabilistic approach to intermittency. *Ergodic Theory Dyn Syst* 19(3):671–685. MR1695915 (2000d:37029)
- Lyons R (1988) On measures simultaneously 2- and 3-invariant. *Israel J Math* 61(2):219–224. MR941238 (89e:28031)
- Liverani C, Wojtkowski MP (1995) Ergodicity in Hamiltonian systems, Dynamics reported. *Dynam. Report. Expositions Dynam. Systems (N.S.)*, vol. 4, Springer, Berlin, pp. 130–202. MR1346498 (96g:58144)
- Manneville P, Pomeau Y (1980) Different ways to turbulence in dissipative dynamical systems. *Phys D* 1(2): 219–226. MR581352 (81h:58041)
- Mañé R (1988) A proof of the $C(X, \mathcal{B}, \mu)$ stability conjecture. *Inst Hautes Études Sci Publ Math* 66: 161–210. MR932138 (89e:58090)
- Mañé R (1987a) Ergodic theory and differentiable dynamics. *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*, vol. 8, Springer-Verlag, Berlin, Translated from the Portuguese by Silvio Levy. MR889254 (88c:58040)

- Marcus B (1978) The horocycle flow is mixing of all degrees. *Invent Math* 46(3):201–209. MR0488168 (58 #7731)
- Marcus B (1975) Unique ergodicity of the horocycle flow: variable negative curvature case. *Israel J Math* 21(2–3):133–144. Conference on Ergodic Theory and Topological Dynamics (Kibbutz Lavi, 1974). MR0407902 (53 #11672)
- Masur H (1986) Closed trajectories for quadratic differentials with an application to billiards. *Duke Math J* 53(2):307–314. MR850537 (87j:30107)
- Mayer DH (1991) Continued fractions and related transformations. *Ergodic theory, symbolic dynamics, and hyperbolic spaces* (Trieste, 1989), Oxford Sci. Publ., Oxford University Press, New York, pp. 175–222. MR1130177
- Melbourne I, Török A (2004) Statistical limit theorems for suspension flows. *Israel J Math* 144:191–209. MR2121540 (2006c:37005)
- Moser J (1962) On invariant curves of area-preserving mappings of an annulus. *Nachr Akad Wiss Göttingen Math-Phys Kl II* 1962:1–20. MR0147741 (26 #5255)
- Méla X, Petersen K (2005) Dynamical properties of the Pascal adic transformation. *Ergodic Theory Dyn Syst* 25(1):227–256. MR2122921 (2005k:37012)
- Nadkarni MG (1998a) Spectral theory of dynamical systems. *Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]*, Birkhäuser Verlag, Basel. MR1719722 (2001d:37001)
- Ornstein D (1970) Bernoulli shifts with the same entropy are isomorphic. *Adv Math* 4:337–352. MR0257322 (41 #1973)
- Ornstein DS, Weiss B (1973) Geodesic flows are Bernoullian. *Israel J Math* 14:184–198. MR0325926 (48 #4272)
- Oseledec VI (1968) A multiplicative ergodic theorem. Characteristic Ljapunov exponents of dynamical systems. *Trudy Moskov Mat Obšč* 19:179–210. (Russian). MR0240280 (39 #1629)
- Parry W (1966) Symbolic dynamics and transformations of the unit interval. *Trans Am Math Soc* 122:368–378. MR0197683 (33 #5846)
- Parry W (1996) Squaring and cubing the circle—Rudolph’s theorem. *Ergodic theory of $Z(X, \mathcal{B}, \mu)$ actions* (Warwick, 1993), London Math Soc Lecture Note Ser, vol. 228, Cambridge University Press, Cambridge, pp. 177–183. MR1411219 (97h:28009)
- Parthasarathy KR (2005) Probability measures on metric spaces. *AMS Chelsea Publishing*, Providence. Reprint of the 1967 original. MR2169627 (2006d:60004)
- Pesin JB (1976) Families of invariant manifolds that correspond to nonzero characteristic exponents. *Izv Akad Nauk SSSR Ser Mat* 40(6):1332–1379, 1440 (Russian). MR0458490 (56 #16690)
- Pesin JB (1977) Characteristic Ljapunov exponents, and smooth ergodic theory. *Uspehi Mat Nauk* 32 no. 4 (196), 55–112, 287 (Russian). MR0466791 (57 #6667)
- Phillips E, Varadhan S (eds) (1975) *Ergodic theory*. Courant Institute of Mathematical Sciences New York University, New York. A seminar held at the Courant Institute of Mathematical Sciences, New York University, New York, 1973–1974; With contributions by S. Varadhan, E. Phillips, S. Alpern, N. Bitzenhofer and R. Adler. MR0486431 (58 #6177)
- Pugh C, Shub M (2004) Stable ergodicity. *Bull Am Math Soc (N.S.)* 41(1):1–41 (electronic). With an appendix by Alexander Starkov. MR2015448 (2005f:37011)
- Robbin JW (1971) A structural stability theorem. *Ann Math* 94:447–493. MR0287580 (44 #4783)
- Robinson C (1973) $C(X, \mathcal{B}, \mu)$ structural stability implies Kupka-Smale. *Dynamical systems (Proc Sympos, Univ Bahia, Salvador, 1971)*. Academic Press, New York, pp. 443–449. MR0334282 (48 #12601)
- Robinson C (1975) Errata to: “Structural stability of vector fields” (*Ann Math* (2) 99 (1974), 154–175). *Ann Math* 101:368. MR0365630 (51 #1882)
- Robinson C (1976) Structural stability of $C(X, \mathcal{B}, \mu)$ diffeomorphisms. *J Differ Equ* 22(1):28–73. MR0474411 (57 #14051)
- Rohlin VA (1952) On the fundamental ideas of measure theory. *Am Math Soc Transl* 1952(71):55. MR0047744 (13,924e)
- Rudolph DJ (1990a) $\times 2$ and $\times 3$ invariant measures and entropy. *Ergodic Theory Dyn Syst* 10(2):395–406. MR1062766 (91g:28026)
- Rudolph DJ (1990b) *Fundamentals of measurable dynamics*. Oxford Science Publications, The Clarendon Press Oxford University Press, New York. *Ergodic theory on Lebesgue spaces*. MR1086631 (92e:28006)
- Ruelle D (1978) Thermodynamic formalism, *Encyclopedia of Mathematics and its Applications*, vol. 5, Addison-Wesley Publishing Co., Reading. The mathematical structures of classical equilibrium statistical mechanics; With a foreword by Giovanni Gallavotti and Gian-Carlo Rota. MR511655 (80g:82017)
- Ruelle D (1976) A measure associated with axiom-A attractors. *Am J Math* 98(3):619–654. MR0415683 (54 #3763)
- Sarig O (2002) Subexponential decay of correlations. *Invent Math* 150(3):629–653. MR1946554 (2004e:37010)
- Schweiger F (1995a) *Ergodic theory of fibred systems and metric number theory*. Oxford Science Publications, The Clarendon Press Oxford University Press, New York. MR1419320 (97h:11083)
- Sinai JG (1970) Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. *Uspehi Mat Nauk* 25(2):141–192 (Russian). MR0274721 (43#481)
- Sinai JG (1968) Markov partitions and U-diffeomorphisms. *Funkcional Anal i Priložen* 2(1):64–89. (Russian). MR0233038 (38 #1361)
- Sinai JG (1972) Gibbs measures in ergodic theory. *Uspehi Mat Nauk* 27(4):21–64 (Russian). MR0399421 (53 #3265)

- Smale S (1980) The mathematics of time. Springer-Verlag, New York. Essays on dynamical systems, economic processes, and related topics. MR607330 (83a:01068)
- Tabachnikov S (2005) Geometry and billiards, Student Mathematical Library, vol 30. American Mathematical Society, Providence. MR2168892 (2006h:51001)
- Thouvenot JP (1995a) Some properties and applications of joinings in ergodic theory. Ergodic theory and its connections with harmonic analysis (Alexandria, 1993), London Math Soc Lecture Note Ser, vol. 205. Cambridge University Press, Cambridge, pp. 207–235. MR1325699 (96d:28017)
- Vershik AM, Livshits AN (1992) Adic models of ergodic transformations, spectral theory, substitutions, and related topics, Representation theory and dynamical systems. Adv Soviet Math, vol. 9, American Mathematical Society, Providence, pp. 185–204. MR1166202 (93i:46131)
- Vorobets YB, Gal'perin GA, Stěpin AM (1992) Periodic billiard trajectories in polygons: generation mechanisms. Uspekhi Mat. Nauk 47(285):9–74, 207 (Russian, with Russian summary); English transl, Russian Math Surveys 47(3) (1992):5–80. MR1185299 (93h:58088)
- Walters P (1982a) An Introduction to Ergodic Theory, Graduate Texts in Mathematics, vol 79. Springer-Verlag, New York. MR648108 (84e:28017)
- Young L-S (1998) Statistical properties of dynamical systems with some hyperbolicity. Ann Math 147(3): 585–650. MR1637655 (99h:58140)
- Young L-S (1999) Recurrence times and rates of mixing. Israel J Math 110:153–188. MR1750438 (2001j:37062)
- Young LS (1993) Ergodic theory of chaotic dynamical systems. In: From Topology to Computation: Proceedings of the Smalefest (Berkeley, CA, 1990), Springer, New York, pp. 201–226. MR1246120 (94i:58112)
- Young L-S (2017) Generalizations of SRB measures to nonautonomous, random, and infinite dimensional systems. J Stat Phys 166(3–4):494–515. <https://doi.org/10.1007/s10955-016-1639-0>. MR3607578
- Zemljakov AN, Katok AB (1975) Topological transitivity of billiards in polygons. Mat Zametki 18(2):291–300. (Russian). MR0399423 (53 #3267)
- Books and Reviews**
- Baladi V (2000b) Positive transfer operators and decay of correlations, Advanced Series in Nonlinear Dynamics, vol 16. World Scientific Publishing Co. Inc., River Edge. MR1793194 (2001k:37035)
- Billingsley P (1978b) Ergodic theory and information. Robert E. Krieger Publishing Co., Huntington. Reprint of the 1965 original. MR524567 (80b:28017)
- Billingsley P (1995) Probability and measure, 3rd ed., Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York. A Wiley-Interscience Publication. MR1324786 (95k:60001)
- Bonatti C, Díaz LJ, Viana M (2005) Dynamics beyond uniform hyperbolicity, Encyclopaedia of Mathematical Sciences, vol. 102, Springer-Verlag, Berlin. A global geometric and probabilistic perspective; Mathematical Physics, III. MR2105774 (2005g:37001)
- Boyarsky A, Góra PL (1997b) Laws of chaos, Probability and its Applications. Birkhäuser Boston Inc., Boston. Invariant measures and dynamical systems in one dimension. MR1461536 (99a:58102)
- Brin M, Stuck G (2002) Introduction to dynamical systems. Cambridge University Press, Cambridge. MR1963683 (2003m:37001)
- Carleson L, Gamelin TW (1993b) Complex dynamics, Universitext: Tracts in Mathematics. Springer-Verlag, New York. MR1230383 (94h:30033)
- Chernov N, Markarian R (2006b) Chaotic billiards, Mathematical Surveys and Monographs, vol 127. American Mathematical Society, Providence. MR2229799 (2007f:37050)
- Collet P, Eckmann J-P (1980b) Iterated maps on the interval as dynamical systems, Progress in Physics, vol 1. Birkhäuser, Boston. MR613981 (82j:58078)
- Cornfeld IP, Fomin SV, Sinai YG (1982b) Ergodic Theory. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 245, Springer-Verlag, New York. Translated from the Russian by A. B. Sosinskii. MR832433 (87f:28019)
- Denker M, Grillenberger C, Sigmund K (1976) Ergodic theory on compact spaces. Springer-Verlag, Berlin. Lecture Notes in Mathematics, vol. 527. MR0457675 (56 #15879)
- Friedman NA (1970) Introduction to Ergodic Theory. Van Nostrand Reinhold Co., New York. Van Nostrand Reinhold Mathematical Studies, No. 29. MR0435350 (55 #8310)
- Glasner E (2003b) Ergodic Theory via Joinings, Mathematical Surveys and Monographs, vol 101. American Mathematical Society, Providence. MR1958753 (2004c:37011)
- Halmos PR (1960) Lectures on Ergodic Theory. Chelsea Publishing Co., New York. MR011817 (22 #2677)
- Hasselblatt B, Katok A (2003b) A first course in dynamics. Cambridge University Press, New York. With a panorama of recent developments. MR1995704 (2004f:37001)
- Hopf E (1937) *Ergodentheorie*, 1st ed., Ergebnisse der Mathematik und ihrer Grenzgebiete 5. Bd., 2. Jaden Springer, Berlin
- Jacobs K (1965) Einige neuere Ergebnisse der Ergodentheorie. Jber Deutsch Math-Verein 67(Abt. 1):143–182 (German). MR0186789 (32 #4244)
- Katok A, Hasselblatt B (1995) Introduction to the modern theory of dynamical systems. In: Encyclopedia of mathematics and its applications, vol. 54, Cambridge University Press, Cambridge. With a supplementary chapter by Katok and Leonardo Mendoza. MR1326374 (96c:58055)
- Keller G (1998) Equilibrium States in Ergodic Theory, London Mathematical Society Student Texts, vol 42. Cambridge University Press, Cambridge. MR1618769 (99e:28022)

- Lucarini V, Faranda D, Moreira ACGM, de Freitas J, Milhazes M, de Freitas M, Holland TK, Nicol M, Todd M, Vaienti S (eds) (2016) *Extremes and recurrence in dynamical systems*, Pure and Applied Mathematics (Hoboken). John Wiley & Sons, Inc., Hoboken. MR3558780
- Mañé R (1987b) *Ergodic Theory and Differentiable Dynamics*. *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*, vol. 8, Springer-Verlag, Berlin. Translated from the Portuguese by Silvio Levy. MR889254 (88c:58040)
- Nadkarni MG (1998b) *Spectral theory of Dynamical Systems*. Birkhäuser Advanced Texts: Basler Lehrbücher [Birkhäuser advanced texts: Basel textbooks]. Birkhäuser Verlag, Basel. MR1719722 (2001d:37001)
- Ornstein DS, Rudolph DJ, Weiss B (1982) Equivalence of measure preserving transformations. *Mem Am Math Soc* 37(262):xii+116. MR653094 (85e:28026)
- Parry W, Pollicott M (1990) Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque* 187–188:268 (English, with French summary). MR1085356 (92f:58141)
- Petersen K (1989) *Ergodic theory*, Cambridge Studies in Advanced Mathematics, vol 2. Cambridge University Press, Cambridge. Corrected reprint of the 1983 original. MR1073173 (92c:28010)
- Royden HL (1988) *Real analysis*, 3rd edn. Macmillan Publishing Company, New York. MR1013117 (90g:00004)
- Rudolph DJ (1990c) *Fundamentals of measurable dynamics*. Oxford Science Publications, The Clarendon Press Oxford University Press, New York, *Ergodic theory on Lebesgue spaces*. MR1086631 (92e:28006)
- Schweiger F (1995b) *Ergodic theory of fibred systems and metric number theory*. Oxford Science Publications, The Clarendon Press Oxford University Press, New York. MR1419320 (97h:11083)
- Thouvenot J-P (1995b) Some properties and applications of joinings in ergodic theory. *Ergodic theory and its connections with harmonic analysis* (Alexandria, 1993), London Math Soc Lecture Note Ser, vol. 205, Cambridge University Press, Cambridge, pp. 207–235. MR1325699 (96d:28017)
- Walters P (1982b) *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics, vol 79. Springer-Verlag, New York. MR648108 (84e:28017)

Ergodicity and Mixing Properties

Terrence Adams¹ and Anthony Quas²

¹Department of Mathematics and Statistics, State University of New York, Albany, NY, USA

²Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada

Article Outline

Glossary
Introduction
Ergodicity
Mixing
Hyperbolicity and Decay of Correlations
Representations, Realizations, and Genericity
Future Directions
References

Ergodicity is the condition under which the strong law of large numbers holds for a dynamical system. A case can be made that ergodicity holds in situations much broader than is typically utilized in statistics, machine learning, and other scientific disciplines. This is not surprising given the power of the pointwise ergodic theorem and the ability to decompose a space into ergodic components. Also, in the first half of the twentieth century, it was shown that topologically generic dynamical systems are weak mixing, rigid, and thus, zero-entropy (and not i.i.d.). This chapter will give a comprehensive account of mixing properties in ergodic theory including Bernoulli, strong mixing, weak mixing, mild, partial, and light mixing. A hierarchy of mixing properties is presented (Fig. 2), along with important examples, as well as applications using the various properties. This chapter places many of the breakthroughs from ergodic theory into a common framework with an eye toward emerging mathematical and scientific areas and concludes with a collection of important unsolved problems.

Glossary

Bernoulli shift Mathematical abstraction of the scenario in statistics or probability in which one performs repeated independent identical experiments.

Markov chain A probability model describing a sequence of observations made at regularly spaced time intervals such that at each time, the probability distribution of the subsequent observation depends only on the current observation and not on prior observations.

Measure-preserving transformation A map from a measure space to itself such that for each measurable subset of the space, it has the same measure as its inverse image under the map.

Measure-theoretic entropy Non-negative (possibly infinite) real number describing the complexity of a measure-preserving transformation.

Product transformation Given a pair of measure-preserving transformations: T of X and S of Y , the product transformation is the map of $X \times Y$ given by $(T \times S)(x, y) = (T(x), S(y))$.

Strong mixing Given two sets, over time the probability that points from one set end up in the other set, converges to the product of the probabilities of the two sets.

Weak mixing There exists a sequence of natural numbers with density 1 in the set of all natural numbers such that the transformation is strong mixing on that sequence.

Introduction

The term “ergodic” was introduced by Boltzmann (1871, 1909) in his work on statistical mechanics, where he was studying Hamiltonian systems with large numbers of particles. The system is described at any time by a point of *phase space*, a subset of $\mathbb{R}^{6N}$ where N is the number of particles. The configuration describes the three-dimensional position and velocity of each of the N particles. It has long

been known that the Hamiltonian (i.e., the overall energy of the system) is invariant over time in these systems. Thus, given a starting configuration, all future configurations as the system evolves lie on the same *energy surface* as the initial one.

Boltzmann's *ergodic hypothesis* was that the trajectory of the configuration in phase space would fill out the entire energy surface. The term "ergodic" is thus an amalgamation of the Greek words for work and path. This hypothesis then allowed Boltzmann to conclude that the long-term average of a quantity as the system evolves would be equal to its average value over the phase space.

Subsequently, it was realized that this hypothesis is rarely satisfied. The ergodic hypothesis was replaced in 1911 by the *quasi-ergodic hypothesis* of the Ehrenfests (1911) which stated instead that each trajectory is dense in the energy surface, rather than filling out the entire energy surface. The modern notion of ergodicity (to be defined below) is due to Birkhoff and Smith (1924). Koopman (1931) suggested studying a measure-preserving transformation by means of the associated isometry on Hilbert space, $U_T: L^2(X) \rightarrow L^2(X)$ defined by $U_T(f) = f \circ T$. This point of view was used by von Neumann (1932) in his proof of the mean ergodic theorem. This was followed closely by Birkhoff (1931) proving the pointwise ergodic theorem. An ergodic measure-preserving transformation enjoys the property that Boltzmann first intended to deduce from his hypothesis: That long-term averages of an observable quantity coincide with the integral of that quantity over the phase space.

These theorems allow one to deduce a form of independence on the average: Given two sets of configurations A and B , one can consider the volume of the phase space consisting of points that are in A at time 0 and in B at time t . In an ergodic measure-preserving transformation, if one computes the average of the volumes of these regions over time, the ergodic theorems mentioned above allow one to deduce that the limit is simply the product of the volume of A and the volume of B . This is the weakest mixing-type property. This chapter outlines a rather full range of mixing properties with ergodicity at the weakest end and the Bernoulli property at the strongest end.

This chapter sets out in some detail the various mixing properties, basing the study on a number

of concrete examples sitting at various points of this hierarchy. Many of the mixing properties may be characterized in terms of the Koopman operators mentioned above (i.e., they are *spectral properties*), but the strongest mixing properties are not spectral in nature. Also, this chapter highlights many of the connections between mixing properties and their spectral reformulations; however, it is recommended the interested reader refer to the chapter ► "[Spectral Theory of Dynamical Systems](#)" (Lemanczyk and Kanigowski 2022) for detailed definitions and results focused mainly on the spectrum of measure preserving systems.

This chapter will also bring to light some of the connections between the range of mixing properties and measure-theoretic entropy. In measure-preserving transformations that arise in practice, there is a correlation between strong mixing properties and positive entropy, although many of these properties are logically independent.

One important issue for which many questions remain open is that of higher-order mixing. Here, the issue is if instead of asking that the observations at two times separated by a large time T be approximately independent, one asks whether if one makes observations at more times, each pair suitably separated, the results can be expected to be approximately independent. This issue has an analogue in probability theory, where it is well-known that it is possible to have a collection of random variables that are pairwise independent, but not mutually independent.

Basics, Examples, and Highlighted Applications

In this chapter, except where otherwise stated, the measure-preserving transformations under consideration are defined on probability spaces.

More specifically, given a measurable space $(X, \mathcal{B})$ and a probability measure μ defined on $\mathcal{B}$, a *measure-preserving transformation* of $(X, \mathcal{B}, \mu)$ is a $\mathcal{B}$ -measurable map $T: X \rightarrow X$ such that $\mu(T^{-1}B) = \mu(B)$ for all $B \in \mathcal{B}$.

While this definition makes sense for arbitrary measures, not simply probability measures, most of the results and definitions below only make sense in the probability measure case. Sometimes it will be helpful to make the assumption that the underlying probability space is a Lebesgue space (i.e., the space

together with its completed σ -algebra agrees up to a measure-preserving bijection a.e. with the unit interval with Lebesgue measure and the usual σ -algebra of Lebesgue measurable sets). Although this sounds like a strong restriction, in practice it is barely a restriction at all, as almost all of the spaces that appear in the theory (and all of those that appear in this chapter) turn out to be Lebesgue spaces. For a detailed treatment of the theory of Lebesgue spaces, the reader is referred to Rudolph's book (1990). The reader is referred also to the chapter on "Measure Preserving Systems".

While many of the definitions presented are valid for both invertible and noninvertible measure-preserving transformations, the strongest mixing conditions are most useful in the case of invertible transformations.

It will be helpful to present a selection of foundational examples, relative to which ergodicity and the various notions of mixing are explored. These examples and the lemmas necessary to show that they are measure-preserving transformations as claimed may be found in the books of Petersen (1983), Rudolph (1990), Walters (1982), and Silva (2008). More details on these examples can also be found in the chapter on ► "Ergodic Theory: Basic Examples and Constructions."

Some Iconic Examples

Example 1 (Rotation on the circle). Let $\alpha \in \mathbb{R}$. Let $R_\alpha : [0, 1) \rightarrow [0, 1)$ be defined by $R_\alpha(x) = x + \alpha \bmod 1$. It is straightforward to verify that R_α preserves the restriction of Lebesgue measure λ to $[0, 1)$ (it is sufficient to check that $\lambda(R_\alpha^{-1}(J)) = \lambda(J)$ for an interval J).

Example 2 (Doubling Map). Let $M_2 : [0, 1) \rightarrow [0, 1)$ be defined by $M_2(x) = 2x \bmod 1$. Again, Lebesgue measure is invariant under M_2 (to see this, one observes that for an interval J , $M_2^{-1}(J)$ consists of two intervals, each of half the length of J). This may be generalized in the obvious way to a map M_k for any integer $k \geq 2$.

Example 3 (Bernoulli Shift). Let A be a finite set and fix a vector $(p_i)_{i \in A}$ of positive numbers that sum to 1. Let $A^{\mathbb{N}}$ denote the set of sequences of the form $x_0x_1x_2 \dots$, where $x_n \in A$ for each $n \in \mathbb{N}$ and let $A^{\mathbb{Z}}$ denote the set of bi-infinite sequences of the

form $\dots x_{-2}x_{-1} \cdot x_0x_1x_2 \dots$ (the $\cdot$ is a placeholder that allows us to distinguish (for example) between the sequences $\dots 01010 \cdot 10101 \dots$ and $\dots 10101 \cdot 01010 \dots$).

A Bernoulli shift is defined as a map (the *shift map*) S on $A^{\mathbb{N}}$ by $(S(x))_n = x_{n+1}$ and define S on $A^{\mathbb{Z}}$ by the same formula. Note that S is invertible as a transformation on $A^{\mathbb{Z}}$ but noninvertible as a transformation on $A^{\mathbb{N}}$.

It is necessary to equip $A^{\mathbb{N}}$ and $A^{\mathbb{Z}}$ with measures. This is done by defining the measure of a preferred class of sets, checking certain consistency conditions and appealing to the Kolmogorov extension theorem. Here the preferred sets are the *cylinder sets*. Given $m \leq n$ in the invertible case and a sequence $a_m \dots a_n$, let $[a_m \dots a_n]_m^n$ denote $\{x \in A^{\mathbb{Z}} : x_m = a_m, \dots, x_n = a_n\}$ and define $\mu([a_m \dots a_n]_m^n) = P_{a_m}P_{a_{m+1}} \dots P_{a_n}$. This is then shown to uniquely define a measure μ on the σ -algebra of $A^{\mathbb{Z}}$ generated by the cylinder sets. It is immediate to see that for any cylinder set C , $\mu(S^{-1}C) = \mu(C)$, and it follows that S is a measure-preserving transformation of $(A^{\mathbb{Z}}, \mathcal{B}, \mu)$. The construction is exactly analogous in the noninvertible case. See the chapter on "Measure Preserving Systems" or the books of Walters (1982) or Rudolph (1990) for more details of defining measures in these systems.

Example 4 (Markov Shift). The spaces $A^{\mathbb{N}}$ and $A^{\mathbb{Z}}$ are exactly as above, as is the shift map. All that changes is the measure.

To define a Markov shift, it is necessary to have a stochastic matrix P (i.e., a matrix with non-negative entries whose rows sum to 1) with rows and columns indexed by A and a left eigenvector π for P with eigenvalue 1 with the property that the entries of π are non-negative and sum to 1. The existence of such an eigenvector is a consequence of the Perron-Frobenius theory of positive matrices. Provided that the matrix P is irreducible (for each a and a' in A , there is an $n > 0$ such that $P^n_{a,a'} > 0$), the eigenvector π is unique.

Given the pair (P, π) , one defines the measure of a cylinder set by $\mu([a_m \dots a_n]_m^n) = \pi_{a_m}P_{a_m a_{m+1}} \dots P_{a_{n-1} a_n}$ and extends μ as before to a probability measure on $A^{\mathbb{N}}$ or $A^{\mathbb{Z}}$.

Example 5 (Interval Exchange Transformation). The class of interval exchange transformations was introduced by Sinai (1976). An interval exchange transformation is the map obtained by cutting the interval into a finite number of pieces and permuting them in such a way that the resulting map is invertible, and restricted to each interval is an order-preserving isometry.

More formally, one takes a sequence of positive lengths $\ell_1, \ell_2, \dots, \ell_k$ summing to 1 and a permutation π of $\{1, \dots, k\}$. For $1 \leq i \leq k$, define $a_i = \sum_{j < i} \ell_j$ and $b_i = \sum_{\pi(j) < \pi(i)} \ell_j$. Note $a_1 = 0$ and set $a_{k+1} = 1$. The interval exchange transformation defined by $(\ell_1, \dots, \ell_k)$ and π is the map $T : [0, 1) \rightarrow [0, 1)$ defined by $T|_{[a_i, a_{i+1})}(x) = x + (b_i - a_i)$. It is straightforward to check that any such interval exchange transformation preserves Lebesgue measure on the unit interval.

Example 6 (Rank-one). There are multiple definitions in the literature for rank-one transformations. For a comprehensive treatment of rank-one or more generally finite rank, see Friedman (1992, 1997); Adams et al. (2017); and Gao and Hill (2014), and for the infinite measure case, see Bozgan et al. (2015).

Now, we give a description of geometric rank-one transformations using the cutting and stacking notation (Friedman 1992; Bozgan et al. 2015). A **column** is a finite sequence of intervals of the same length, sometimes also called **levels**, that are usually taken left-closed and right-open; its **height** is the number of intervals in the column. A column C defines a column map T_C so that each interval is sent to the interval right above it by translation; so T_C is defined on every interval of the column except the top. It is clear that on the intervals where it is defined, T_C is measurable and measure-preserving with respect to Lebesgue measure. To specify a rank-one construction, a sequence of positive integers $(r_n)_{n \geq 0}$, $r_n \geq 2$, and a doubly indexed sequence of nonnegative integers $(s_{n,i})$, $n \geq 0$, $i = 0, \dots, r_n - 1$ are given. Begin with column C_0 consisting of a single finite interval, so its height is $h_0 = 1$. For the inductive step, suppose column C_n of height h_n are given. Then cut each interval of column C_n into r_n subintervals

of equal length to obtain r_n subcolumns of equal measure. For each $i = 0, \dots, r_n - 1$, place $s_{n,i}$ new intervals, called **spacers**, above the i th subcolumn. Then C_{n+1} is constructed by stacking each subcolumn, right over left, with its associated spacers under the next subcolumn. The example below (Fig. 1) shows this in more detail. Thus C_{n+1} will consist of r_n copies of C_n possibly separated by spacers. From the construction, it is clear that $T_{C_{n+1}}$ agrees with T_{C_n} wherever T_{C_n} is defined. Let X be the union of the levels in the columns and $T(x) = \lim_{n \rightarrow \infty} T_{C_n}$. The height of $T_{C_{n+1}}$ is given by

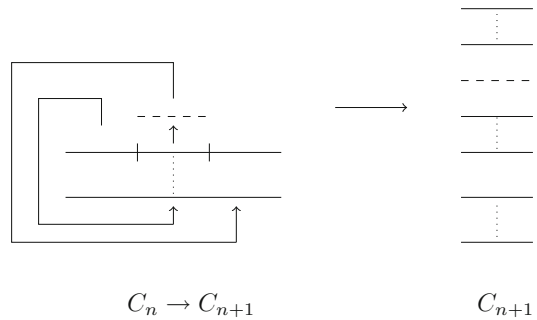
$$h_{n+1} = r_n h_n + \sum_{i=0}^{r_n-1} s_{n,i}.$$

The transformation T will be finite measure-preserving, if

$$\sum_{n=0}^{\infty} \frac{\sum_{i=0}^{r_n-1} s_{n,i}}{h_{n+1}} < \infty,$$

otherwise, T will be infinite measure-preserving. In the case that T is finite measure-preserving, we can choose the length of the original interval C_0 such that $X = [0, 1)$ a.e. and T is an invertible Lebesgue probability preserving transformation on X .

Example 7 (The Chacon Transformations). Constructions of measure-preserving transformations



Ergodicity and Mixing Properties, Fig. 1 Construction of the Chacon-3 transformation

satisfying certain properties are not always easy to come by. In [68], it is shown that generically in the weak topology, invertible measure-preserving transformations are weak mixing and rigid (which implies not strong mixing). However, for some years, there was a dearth of explicit constructions that are weak mixing and not strong mixing (Chacon 1966a, b; 1969; Maruyama 1949).¹ See Sect. 4 for a more detailed description of strong and weak mixing, and Sect. 6.1 for a description of the weak topology on invertible measure-preserving transformations.

The rank-one transformations (Chacon 1966b; 1969), first published by Chacon as weak mixing and not strong mixing, are not the same as the transformation commonly referred to as Chacon's transformation (Friedman 1970, p. 86). While these transformations are all weak mixing, the transformation commonly referred to as Chacon's transformation uses a cut into three subcolumns at each stage with a single spacer added on the second subcolumn (i.e., $r_n = 3$, $s_{n,0} = 0$, $s_{n,1} = 1$, $s_{n,2} = 0$, $h_{n+1} = 3h_n + 1$). It is depicted in Fig. 1. This transformation is called Chacon-3.

The transformation published in Chacon (1969) is referred to as Chacon-2. It is constructed by cutting each column into two subcolumns at each stage and adding a single spacer on the last subcolumn. Thus, $r_n = 2$, $s_{n,0} = 0$, $s_{n,1} = 1$, $h_n = 2^{n+1} - 1$ for all n .

Factors, Extensions, and Isomorphisms

This chapter makes use of the concept of measure-theoretic homomorphisms, isomorphisms, and factor systems. Let $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{F}, \nu, S)$ be measure-preserving dynamical systems on Lebesgue probability spaces. The system $(Y, \mathcal{F}, \nu, S)$ is a factor of $(X, \mathcal{B}, \mu, T)$, if there exists a measure-preserving transformation ϕ of X onto Y such that for a.e. $x \in X$,

$$\phi(Tx) = S(\phi x).$$

The map ϕ is a measure-preserving homomorphism and $(X, \mathcal{B}, \mu, T)$ is called an extension of $(Y, \mathcal{F}, \nu, S)$. If the map ϕ is a bijection a.e., then ϕ is called an isomorphism, and the systems $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{F}, \nu, S)$ are said to be (measure theoretically) isomorphic. When the underlying probability spaces are clear, it may be said that T and S are isomorphic.

Measure-theoretic isomorphism is the basic notion of "sameness" in ergodic theory. It is in some sense quite weak, so that systems may be isomorphic that feel very different. For example, as discussed later, the time one map of a geodesic flow is isomorphic to a Bernoulli shift). For comparison, the notion of sameness in topological dynamical systems (topological conjugacy) is far stronger.

As an example of measure-theoretic isomorphism, it may be seen that the doubling map is isomorphic to the one-sided Bernoulli shift on $\{0, 1\}$ with $p_0 = p_1 = 1/2$ (the map ϕ takes an $x \in [0, 1)$ to the sequence of 0's and 1's in its binary expansion; choosing the sequence ending with 0's, for example, if x is of the form $p/2^n$).

Given a measure-preserving transformation T of a probability space $(X, \mathcal{B}, \mu)$, T is associated to an isometry of $L^2(X, \mathcal{B}, \mu)$ by $U_T(f) = f \circ T$. This operator is known as the *Koopman Operator*. In the case where T is invertible, the operator U_T is unitary. Two measure-preserving transformations T and S of $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{F}, \nu)$ are *spectrally isomorphic* if there is a Hilbert space isomorphism Θ from $L^2(X, \mathcal{B}, \mu)$ to $L^2(Y, \mathcal{F}, \nu)$ such that $\Theta \circ U_T = U_S \circ \Theta$. As demonstrated below, spectral isomorphism is a strictly weaker property than measure-theoretic isomorphism.

Since in ergodic theory, measure-theoretic isomorphism is the basic notion of sameness, most properties that are used to describe measure-preserving systems are required to be invariant under measure-theoretic isomorphism (i.e., if two measure-preserving transformations are measure-theoretically isomorphic, the first has a given property if and only if the second does). On the other hand, we shall see that some mixing-type properties are invariant under spectral

¹From Parry quoted in Kakutani (1986), "For many years the first of Kakutani's examples (which he had discussed with von Neumann) was the only other known example. This example became widely known through the publicity of Kakutani's lectures. It appears for example in Friedman's book."

isomorphism, while others are not. If a property is invariant under spectral isomorphism, it is said that it is a *spectral property*.

There are a number of mixing type properties that occur in the probability literature (α -mixing, β -mixing, ϕ -mixing, ψ -mixing, etc.) (see Bradley's (2005) survey for a description of these conditions). In the case of a stationary process, X_i for $-\infty < i < \infty$, the process can be modeled using a measure preserving transformation T and measurable function f such that $X_i = f \circ T^{i-1}$. These mixing properties from the probability literature often constrain the function f and are not preserved under measure-theoretic isomorphism. One example of this distinction is that finite ergodic Markov chains are measure-theoretically isomorphic to Bernoulli shifts; however, these systems are viewed as distinct varieties of stationary processes. For this reason, these properties are not widely used in ergodic theory, although β -mixing turns out to be equivalent to the so-called *weak Bernoulli* property (which turns out to be stronger than the Bernoulli property that is discussed in this chapter – see Smorodinsky's (1971) paper) and α -mixing is equivalent to strong mixing.

A basic construction (see the chapter on ▶ “Ergodic Theory: Basic Examples and Constructions”) that is required in what follows is the product of a pair of measure-preserving transformations: Given transformations T of $(X, \mathcal{B}, \mu)$ and S of $(Y, \mathcal{F}, \nu)$, the *product transformation* is defined as $T \times S$: $(X \times Y, \mathcal{B} \otimes \mathcal{F}, \mu \times \nu)$ by $(T \times S)(x, y) = (Tx, Sy)$.

One issue that is faced on occasion is that it is sometimes convenient to deal with noninvertible measure-preserving transformations. It turns out that given a noninvertible measure-preserving transformation, there is a natural way to uniquely associate an invertible measure-preserving transformation sharing almost all of the ergodic properties of the original transformation. Specifically, given a noninvertible measure-preserving transformation T of $(X, \mathcal{B}, \mu)$, one lets $\bar{X} = \{(x_0, x_1, \dots) : x_n \in X \text{ and } T(x_n) = x_{n-1} \text{ for all } n\}$, $\bar{\mathcal{B}}$ be the σ -algebra generated by sets of the form $\bar{A}_n = \{\bar{x} \in \bar{X} : x_n \in A\}$, $\bar{\mu}(\bar{A}_n) = \mu(A)$ and

$\bar{T}(x_0, x_1, \dots) = (T(x_0), x_0, x_1, \dots)$. The transformation $\bar{T}$ of $(\bar{X}, \bar{\mathcal{B}}, \bar{\mu})$ is called the *natural extension* of the transformation T of $(X, \mathcal{B}, \mu)$ (see the chapter on “Ergodic Theory: Basic Examples and Constructions” for more details). In situations where one wants to use invertibility, it is often possible to pass to the natural extension, work there, and then derive conclusions about the original noninvertible transformation.

Arguably, the simplest example of an extension is a two-point extension. Given an ergodic measure-preserving transformation T acting on $(X, \mathcal{B}, \mu)$, this can be viewed as selecting a switching set $A \subset X$ of positive measure. The two-point extension is composed of two copies $(X_1, \mathcal{B}, \mu, T_1)$ and $(X_2, \mathcal{B}, \mu, T_2)$ with mirror copies A_1, A_2 of the switching set A . Also, there is a measure-preserving bijection $\phi : X_1 \rightarrow X_2$ between the two mirror copies such that $A_2 = \phi(A_1)$. The two-point extension S is defined as,

$$S(x) = \begin{cases} T_1(x) & \text{if } x \in X_1 \setminus A_1 \\ T_2(x) & \text{if } x \in X_2 \setminus A_2 \\ T_2(\phi(x)) & \text{if } x \in A_1 \\ T_1(\phi^{-1}(x)) & \text{if } x \in A_2. \end{cases}$$

Establishing that mixing properties such as Bernoulli extend to a two-point extension is already nontrivial (Rudolph 1978; Kammeyer 1990). These notions generalize to more general group extensions, as well as nonalgebraic extensions such as the roof over a transformation. A roof over a transformation is constructed by identifying a subset $A \subset X$ of positive measure and extending the measure space using a measurable set A_0 and invertible measure preserving isomorphism ϕ such that $A_0 = \phi(A)$ is disjoint from X . The set A_0 is a disjoint copy of A . In this case, the extension S is defined as,

$$S(x) = \begin{cases} T(x) & \text{if } x \in X \setminus A \\ \phi(x) & \text{if } x \in A \\ T(\phi^{-1}(x)) & \text{if } x \in A_0. \end{cases}$$

The normalized measure μ_0 defined as $\mu_0(B) = \mu(B)/(1 + \mu(A_0))$ is an invariant probability

measure under the action of S . The first weak mixing, nonstrong mixing transformation due to Kakutani and von Neumann was constructed as a carefully designed roof over an irrational rotation (see previous footnote 1).

The notion of induced (or derived) transformation (Kakutani 1943) is the reverse notion from the “roof” extension. Given a subset $A \subset X$ of positive measure, the induced transformation T_A is defined on A as

$$T_A(x) = T^{r_A(x)}(x),$$

where $r_A(x) = \inf \{n \geq 1 : T^n(x) \in A\}$. If T is ergodic, invertible, and measure-preserving, then so is T_A . Also, the entropy of an induced transformation may be computed directly from the entropy $h(T)$ of T , as $h(T_A) = \frac{1}{\mu(A)} h(T)$ (Abramov 1959). Interestingly, it was established for an ergodic transformation, there exists a dense collection of sets A such that T_A is weak mixing (Chacon 1966a), and then a dense collection of A such that T_A is strongly mixing (Friedman and Ornstein 1973). Hence, one method for obtaining a zero-entropy strong mixing transformation is to apply the result of Friedman-Ornstein (1973) to any zero-entropy ergodic transformation. The notion of an induced transformation has led to the theory of Kakutani equivalence and many important results for both transformations and flows. For example, see results of Ambrose and Kakutani (1942) and Rudolph (1976).

Highlighted Applications and Connections

Application 1 (Hard Sphere Gases and Billiards). We wish to model the behavior of a gas in a bounded region. We make the assumption that the gas consists of a large number N of identical balls which move at constant velocity until two balls collide, whereupon they elastically swap momentum along the direction of contact. The phase space for this system is a region of $\mathbb{R}^{6N}$ (with N 3-dimensional position vectors and N 3-dimensional velocity vectors). More abstractly, the system is equivalent to the motion of a single point particle in a region of $\mathbb{R}^M \times \mathbb{R}^M$ (with the first M -vector representing position and the

second representing velocity). The system is constrained in that its position is required to lie in a bounded region S of $\mathbb{R}^M$ with a piecewise smooth boundary. The system evolves by moving the position at a constant rate in the direction of the velocity vector until the point reaches ∂S , at which time the component of the velocity parallel to the normal to ∂S is reversed. This then defines a flow (i.e., a family of maps $(T_t)_{t \in \mathbb{R}}$ satisfying $T_{t+s} = T_t \circ T_s$) on the phase space. Since the magnitude of the velocity is conserved, it is convenient to restrict to flows with speed 1. This system is clearly the closest of the examples that we consider to the situation envisaged by Boltzmann. Perhaps not surprisingly, proofs of even the most basic properties for this system are much harder than the other examples that we consider.

Application 2 (Combinatorial Number Theory).

The legendary mathematician Paul Erdős was known to disavow life's material possessions and instead, give complete devotion to mathematics, and of course, his own mother. He used all disposable cash to fund a growing number of bounties on unsolved math problems. His first major award went to Szemerédi for proving that any sequence of natural numbers with positive upper density contains arbitrarily long arithmetic progressions. The award was \$1000 in 1974. Fast forward to 2004, mathematicians Green and Tao extended Szemerédi's results to prove that the prime numbers contain arbitrarily long arithmetic progressions (Green and Tao 2008). A conjecture of Erdős continues to remain open: Given any increasing sequence of natural numbers, n_i such that $\sum_{i=1}^{\infty} \frac{1}{n_i} = \infty$, then n_i contains arbitrarily long arithmetic progressions. A positive answer to this question would imply the primes contain arbitrarily long arithmetic progressions.

In the meantime (circa 1977), Hillel Furstenberg proved Szemerédi's theorem using ergodic theory techniques (Furstenberg 1977). Furstenberg established a transference principle that recast number theoretic questions into properties of measure-preserving dynamical systems. Furstenberg, Bergelson, and others spearheaded new research on multiple

recurrence of dynamical systems to obtain far reaching generalizations of the original Szemerédi's theorem, including to higher order group actions (Furstenberg and Katznelson 1978; Bergelson and Leibman 2002) and probabilistic versions for occurrences of arithmetic progressions based on multi-dimensional ergodic theorems (Bergelson 1987; Conze and Lesigne 1984; Zhang 1996; Frantzikinakis and Kra 2005; Tao 2008; Austin 2010). Many of the ideas from Furstenberg's ergodic theoretic approach find themselves in the ideas of Green and Tao. To better comprehend the impact of Furstenberg's breakthroughs, as well as Tao and Green's, here is an excerpt from Tao's 2006 Fields Medal citation:

Of special note is his joint work with Ben Green, a Clay Research Fellow from 2005 through 2007. In their 2004 paper, "The primes contain arbitrarily long arithmetic progressions," the authors answered in the affirmative a long-standing conjecture that had resisted many attempts.

Application 3 (Statistical Learning Theory). Statistical learning theory is the application of statistics, probability, and mathematics to improve the understanding of machine learning algorithms and to measure the effectiveness of mathematical models to learn from data, predict new outcomes, and approximate quantitative relationships. Statistical learning theory has experienced growing interest due to renewed interest in machine learning, neural networks, and in particular, deep neural networks. Deep neural networks (DNNs) are achieving new state-of-the-art in many machine learning areas such as natural language processing, speech, and image recognition. However, DNNs are not understood well mathematically, nor is it easy to predict their failure modes in real-life situations.

Statisticians often model sequential data using random variables defined on a probability space. If the random variables are independent and identically distributed, there are many powerful results such as the central limit theorem or uniform convergence of ergodic averages for classes of models with bounded complexity. In particular, Vapnik and Chervonenkis introduced VC-dimension as a measure of complexity for a collection of sets and then showed that pointwise ergodic averages converge uniformly over a class of sets with bounded

VC-dimension for almost all i.i.d. samples (Vapnik and Chervonenkis 2015).

Using ergodic theory techniques, it was shown that Vapnik-Chervonenkis uniform convergence holds for stationary ergodic processes (Adams and Nobel 2010).

Theorem 1 (Adams and Nobel 2010). *Let χ be a complete separable metric space equipped with its Borel measurable subsets S , and let $C \subset S$ be any countable family of sets. If $\dim(C) < \infty$, then for every stationary ergodic process $\mathbf{X} = X_1, X_2, \dots$ taking values in (χ, S) ,*

$$\Gamma_m(C : \mathbf{X}) = \sup_{C \in \mathcal{C}} \left| \frac{1}{m} \sum_{i=1}^m I(X_i \in C) - \mathcal{P}(X \in C) \right| \rightarrow 0 \text{ wpl} \quad (1)$$

as m tends to infinity. In other words, $\mathcal{C}$ is a Glivenko-Cantelli class for every stationary ergodic process.

In ergodic theory notation, this result implies for a collection $\mathcal{C}$ with finite VC-dimension and a Lebesgue probability space $(X, \mathcal{B}, \mu)$, for almost every $x \in X$,

$$\lim_{n \rightarrow \infty} \sup_{C \in \mathcal{C}} \left| \frac{1}{n} \sum_{i=0}^{n-1} I_C(T^i x) - \mu(C) \right| = 0.$$

Note that ergodic stationary processes can be modeled using an ergodic invertible measure-preserving transformation T and measurable function f such that $X_i(x) = f(T^{i-1}x)$. It is possible to extend this result to processes of the form $f(T^i x)$ if the complexity of the class of functions that f is drawn from is restricted. See Vapnik (2013); Pollard (1990); Adams and Nobel (2010) for further details.

Application 4 (Consistent Temporal Modeling by HMMs and DNNs). Hidden Markov Models (HMMs) have been studied for decades and were commonly used for modeling data from various domains (i.e., text, speech, video). The power of HMMs came from the ability to develop efficient

algorithms for approximately learning transition and emission probabilities. HMMs have been studied in the context of ergodic theory under different frameworks (e.g., subshifts of finite type, sofic measures). See Marcus et al. (2011) for more details on applications of ergodic theory and entropy to HMMs. In McGoff et al. (2015), the consistency of maximum likelihood estimates for HMMs was established under general conditions including observations with noise.

More recently, the availability of big data has led to a resurgence in neural networks and in particular, deep neural networks (DNNs) for data modeling and prediction. While DNNs may be successful in modeling data (examples include Google translate, Apple Siri, Amazon Alexa), the DNNs are generally viewed as black boxes with little understanding of why and when they work. This lack of understanding has spurred new research toward applying rigorous mathematics to improve understanding of deep neural networks, both in the training phase and in the understanding of failure modes after a system is trained. During 2018–2019, several research papers have demonstrated how over-parameterization of neural networks can lead to consistent modeling. A key ingredient in this approach is the use of Hoeffding’s inequality to bound the probability that the sum of bounded independent random variables deviates from its expected value by more than a certain amount. This is a computational version of the ergodic theorem which does not hold in all situations. However, there is research to extend Hoeffding’s inequality beyond the i.i.d. case (Impagliazzo and Kabanets 2010; Schmidt et al. 1993; Pelekis and Ramon 2017), in some cases developing new concentration inequalities. For more details on recent research in this area, see Du et al. (2019); Arora et al. (2019); Arora et al. (2018); and Raginsky et al. (2017).

Ergodicity

Given a measure-preserving transformation $T: X \rightarrow X$, if $T^{-1}A = A$, then $T^{-1}A^c = A^c$ also.

This allows us to decompose the transformation X into two pieces A and A^c and study the transformation T separately on each. In fact the same situation holds if $T^{-1}A$ and A agree up to a set of measure 0. For this reason, a set A is called *invariant* if $\mu(T^{-1}A \Delta A) = 0$.

Returning to Boltzmann’s ergodic hypothesis, existence of an invariant set of measure between 0 and 1 would be a bad situation as his essential idea was that the orbit of a single point would “see” all of X , whereas if X were decomposed in this way, the most that a point in A could see would be all of A , and similarly the most that a point in A^c could see would be all of A^c .

A measure-preserving transformation will be called *ergodic* if it has no nontrivial decomposition of this form. More formally, let T be a measure-preserving transformation of a probability space $(X, \mathcal{B}, \mu)$. The transformation T is said to be ergodic if for all invariant sets, either the set or its complement has measure 0.

Unlike the remaining concepts discussed in this chapter, this definition of ergodicity applies also to infinite measure-preserving transformations and even to certain non-measure-preserving transformations. See Aaronson’s book (Aaronson 1997) for more information.

The following lemma is often useful:

Lemma 2 *Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T: X \rightarrow X$ be a measure-preserving transformation. Then T is ergodic if and only if the only measurable functions f satisfying $f \circ T = f$ (up to sets of measure 0) are constant almost everywhere.*

For the straightforward proof, notice that if the condition in the lemma holds and A is an invariant set, then $\mathbf{1}_A \circ T = \mathbf{1}_A$ almost everywhere, so that $\mathbf{1}_A$ is an a.e. constant function and so A or A^c is of measure 0. Conversely, if f is an invariant function, it is seen that for each a , $\{x: f(x) < a\}$ is an invariant set and hence of measure 0 or 1. It follows that f is constant almost everywhere. Note that it is sufficient to check that the bounded measurable invariant functions are constant.

The following corollary of the lemma shows that ergodicity is a spectral property.

Corollary 3 *Let T be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. Then T is ergodic if and only if 1 is a simple eigenvalue of U_T .*

The ergodic theorems mentioned earlier due to von Neumann and Birkhoff are the following (see also the chapter on ► “Ergodic Theorems”).

Theorem 4 (von Neumann Mean Ergodic Theorem (von Neumann 1932)). *Let T be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. For $f \in L^2(X, \mathcal{B}, \mu)$, let $A_N f = (f + f \circ T + \dots + f \circ T^{N-1})/N$. Then for all $f \in L^2(X, \mathcal{B}, \mu)$, $A_N f$ converges in L^2 to an invariant function f^* .*

Theorem 5 (Birkhoff Pointwise Ergodic Theorem (Birkhoff 1931)). *Let T be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. Let $f \in L^1(X, \mathcal{B}, \mu)$. Let $A_N f$ be as above. Then for μ -almost every $x \in X$, $(A_N f(x))$ is a convergent sequence.*

Of these two theorems, the pointwise ergodic theorem is the deeper result, and it is straightforward to deduce the mean ergodic theorem from the pointwise ergodic theorem. The mean ergodic theorem was reproved very concisely by Riesz (1938) and it is this proof that is widely known now. Riesz’s proof is reproduced in Parry’s (1981) book. There have been many different proofs given of the pointwise ergodic theorem. Notable among these are the argument due to Garsia (1965) and a proof due to Katznelson and Weiss (1982) based on work of Kamae (1982), which appears in a simplified form in work of Keane and Petersen (2006).

If the measure-preserving transformation T is ergodic, then by virtue of Lemma 2, the limit functions appearing in the ergodic theorems are constant. One sees that the constant is simply the integral of f with respect to μ , so that in this situation $A_N f(x)$ converges to $\int f d\mu$ in norm and pointwise almost everywhere, thereby providing a justification of Boltzmann’s original claim: For ergodic measure-preserving transformations, time averages agree with spatial averages.

In the case where T is not ergodic, it is also possible to identify the limit in the ergodic theorems: $f^* = \mathbb{E}(f|\mathcal{I})$, where $\mathcal{I}$ is the σ -algebra of T -invariant sets.

Note that the set on which the almost everywhere convergence in Birkhoff’s theorem takes place depends on the L^1 function f that one is considering. Straight-forward considerations show that there is no single full measure set that works simultaneously for all L^1 functions. In the case where X is a compact metric space, it is well known that $C(X)$, the space of continuous functions on X with the uniform norm, has a countable dense set which can be denoted by $(f_n)_{n \geq 1}$. If the dynamical system $(X, \mathcal{B}, \mu, T)$ is measure preserving and ergodic, then for each n , there is a set B_n of measure 1 such that for all $x \in B_n$, $A_N f_n(x) \rightarrow f_n d\mu$. Letting $B = \bigcap_n B_n$, one obtains a full measure set such that for all n and all $x \in B$, $A_N f_n(x) \rightarrow \int f_n d\mu$. A simple approximation argument then shows that for all $x \in B$ and all $f \in C(X)$, $A_N f(x) \rightarrow \int f d\mu$. A point x with this property is said to be *generic* for $(X, \mathcal{B}, \mu, T)$. The observations above show that for an ergodic measure preserving system $(X, \mathcal{B}, \mu, T)$, $\mu\{x: x \text{ is generic for } (X, \mathcal{B}, \mu, T)\} = 1$.

If T is ergodic, but T^n is not ergodic for some n , then one can show that the space X splits up as $A_1, \dots, A_d$ for some $d|n$ in such a way that $T(A_i) = A_{i+1}$ for $i < d$ and $T(A_d) = A_1$ with T^n acting ergodically on each A_i . The transformation T is *totally ergodic* if T^n is ergodic for all $n \in \mathbb{N}$. One can check that a noninvertible transformation T is ergodic if and only if its natural extension is ergodic.

The following lemma gives an alternative characterization of ergodicity, which in particular relates it to mixing.

Lemma 6 (Ergodicity as a Mixing Property). *Let T be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. Then T is ergodic if and only if for all f and g in L^2 ,*

$$\frac{1}{N} \sum_{n=0}^N \langle f, g \circ T^n \rangle \rightarrow \langle f, 1 \rangle \langle 1, g \rangle.$$

In particular, if T is ergodic, then $(1/N)\sum_{n=0}^{N-1}\mu(A \cap T^{-n}B) \rightarrow \mu(A)\mu(B)$ all measurable sets A and B .

Now, this chapter examines the ergodicity of the examples presented above. First, for the rotation of the circle, it is claimed that the transformation is ergodic if and only if the ‘angle’ α is irrational. To see this, argue as follows. If $\alpha = p/q$, then $f(x) = e^{2\pi i q x}$ is a nonconstant R_α -invariant function, and hence R_α is not ergodic. On the other hand, if α is irrational, suppose f is a bounded measurable invariant function. Since f is bounded, it is an L^2 function, and so f may be expressed in L^2 as a Fourier series: $f = \sum_{n \in \mathbb{Z}} c_n e_n$ where $e_n(x) = e^{2\pi i n x}$. It is seen that $f \circ R_\alpha = \sum_{n \in \mathbb{Z}} e^{2\pi i n \alpha} c_n e_n$. In order for f to be equal in L^2 to $f \circ R_\alpha$, they must have the same Fourier coefficients, so that $c_n = e^{2\pi i n \alpha} c_n$ for each n . Since α is irrational, this forces $c_n = 0$ for all $n \neq 0$, so that f is constant as required.

The doubling map and the Bernoulli shift are both ergodic, although we defer proof of this for the time being, since they in fact have the strong mixing property. A Markov chain with matrix P and vector π is ergodic if and only if for all i and j in A with $\pi_i > 0$ and $\pi_j > 0$, there exists an $n \geq 0$ with $P_{ij}^n > 0$. This follows from the ergodic theorem for Markov chains (which is derived from the Strong Law of Large Numbers) (see Feller (1950) for details). In particular, if the underlying Markov chain is irreducible, then the measure is ergodic.

In the case of interval exchange transformations, there is a simple necessary condition on the permutation for irreducibility, namely, for $1 \leq j \leq k$, $\pi\{1, \dots, j\} \neq \{1, \dots, j\}$. Under this condition, Masur (1982) and Veech (1982) independently showed that for almost all values of the sequence of lengths $(\ell_i)_{1 \leq i \leq k}$, the interval exchange transformation is ergodic. (In fact they showed the stronger condition of *unique ergodicity*: That the transformation has no other invariant measure than Lebesgue measure. This implies that Lebesgue measure is ergodic, because if there were a nontrivial invariant set, then the restriction of Lebesgue measure to that set would be another invariant measure.)

To see that any rank-one measure-preserving transformation is ergodic, note that each level sweeps out the entire space under iterates of T . In particular, if I is a level in a column C_n , then $\mu(\cup_{i=0}^{\infty} T^i(I)) = 1$. By approximation, every set A of positive measure satisfies, $\mu(\cup_{i=0}^{\infty} T^i(A)) = 1$. This proves that every rank-one transformation is ergodic, including Chacon-2 and Chacon-3.

For the hard sphere systems, there are no results on ergodicity in full generality. Important special cases have been studied by Sinai (1970), Sinai and Chernov (1987), Krámli et al. (1991), Simányi and Szász (1999), Simányi (2004, 2003), and Young (1998).

Ergodic Decomposition

It has already been observed that if a transformation is not ergodic, then it may be decomposed into parts. Clearly if these parts are not ergodic, they may be further decomposed. It is natural to ask whether the transformation can be decomposed into ergodic parts, and if so what form does the decomposition take? In fact such a decomposition does exist, but rather than decompose the transformation, it is necessary to decompose the measure into ergodic pieces. This is known as ergodic decomposition.

The set of invariant measures for a measurable map T of a measurable space $(X, \mathcal{B})$ to itself forms a simplex. General functional analytic considerations (due to Choquet (1956b, a) – see also Phelps’s (1966) account of this theory) mean that it is possible to write any member of the simplex as an integral-convex combination of the extreme points. Further, the extreme points of the simplex may be identified as precisely the ergodic invariant measures for T . It follows that any invariant probability measure μ for T may be uniquely expressed in the form

$$\mu(A) = \int_{M_{\text{erg}}(X, T)} \nu(A) dm(\nu),$$

where $M_{\text{erg}}(X, T)$ denotes the set of ergodic T -invariant measures on X and m is a measure on $M_{\text{erg}}(X, T)$.

A detailed proof of ergodic decomposition is not given here. The theorem can be proved using the Birkhoff ergodic theorem and the Riesz Representation theorem identifying the dual space of the space of continuous functions on a compact space as the set of bounded signed measures on the space (see Rudin's (1966) book for details). For more details on ergodic decomposition, see Rudolph's (1990) book which gives a full proof in the case that X is a Lebesgue space based on a detailed development of the theory of these spaces and builds measures using conditional expectations. Oxtoby (1952) also wrote a survey article containing many of the details (and much more besides).

Theorem 7 *Let X be a compact metric space, $\mathcal{B}$ be the Borel σ -algebra, μ be an invariant Borel probability measure and T be a continuous measure-preserving transformation of $(X, \mathcal{B}, \mu)$. Then for each $x \in X$, there exists an invariant Borel measure μ_x such that:*

1. For $f \in L^1(X, \mathcal{B}, \mu)$, $\int f d\mu = \int (\int f d\mu_x) d\mu(x)$;
2. Given $f \in L^1(X, \mathcal{B}, \mu)$, for μ -almost every $x \in X$, one has $A_N f(x) \rightarrow \int f d\mu_x$;
3. The measure μ_x is ergodic for μ -almost every $x \in X$.

Notice that conclusion (2) shows that μ_x can be understood as the distribution on the phase space "seen" if one starts the system in an initial condition of x . This interpretation of the measures μ_x corresponds closely with the ideas of Boltzmann and the Ehrenfests in the formulation of the ergodic and quasi-ergodic hypotheses, which can be seen as demanding that μ_x is equal to μ for (almost) all x .

Mixing

As mentioned above, ergodicity may be seen as an independence on average property. More specifically, one wants to know whether in some sense $\mu(A \cap T^{-n}B)$ converges to $\mu(A)\mu(B)$ as $n \rightarrow \infty$. Ergodicity is the property that there is convergence in the Césaro sense. Weak mixing is the

property that there is convergence in the strong Césaro sense. That is, a measure-preserving transformation T is *weak mixing* if

$$\frac{1}{N} \sum_{n=0}^{N-1} |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| \rightarrow 0 \text{ as } N \rightarrow \infty.$$

In order for T to be *strong mixing*, it is required simply $\mu(A \cap T^{-N}B) \rightarrow \mu(A)\mu(B)$ as $N \rightarrow \infty$. It is clear that strong mixing implies weak mixing and weak mixing implies ergodicity.

If T^d is not ergodic (so that $T^{-d}A = A$ for some A of measure strictly between 0 and 1), then $|\mu(T^{-nd}A \cap A) - \mu(A)^2| = \mu(A)(1 - \mu(A))$, so that T is not weak mixing.

An alternative characterization of weak mixing is as follows:

Lemma 8 *The measure-preserving transformation T is weak mixing if and only if for every pair of measurable sets A and B , there exists a subset J of $\mathbb{N}$ of density 1 (i.e., $\#(J \cap \{1, \dots, N\})/N \rightarrow 1$) such that*

$$\lim_{n \rightarrow \infty, n \notin J} \mu(A \cap T^{-n}B) = \mu(A)\mu(B). \quad (2)$$

By taking a countable family of measurable sets that are dense (with respect to the metric $d(A, B) = \mu(A \Delta B)$) and taking a suitable intersection of the corresponding J sets, one shows that for a given weak mixing measure-preserving transformation, there is a single set $J \subset \mathbb{N}$ such that (2) holds for *all* measurable sets A and B (see Petersen (1983) or Walters (1982) for a proof).

It is shown that an irrational rotation of the circle is not weak mixing as follows: let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and let A be the interval $[\frac{1}{4}, \frac{3}{4})$. There is a positive proportion of n 's in the natural numbers (in fact proportion 1/3) with the property that $|T^n(\frac{1}{2}) - \frac{1}{2}| < \frac{1}{6}$. For these n 's $\mu(A \cap T^{-n}A) > \frac{1}{3}$, so that in particular $|\mu(A \cap T^{-n}A) - \mu(A)\mu(A)| > \frac{1}{12}$. Clearly this precludes the required convergence to 0 in the definition of weak mixing. Since $R_\alpha^n = R_{n\alpha}$, the earlier argument shows that R_α^n is ergodic, so that R_α is totally ergodic, but not weak mixing.

On the other hand, now it is shown that any Bernoulli shift is strong mixing. To see this, let A and B be arbitrary measurable sets. By standard measure-theoretic arguments, A and B may each be approximated arbitrarily closely by a finite union of cylinder sets. Since if A' and B' are finite unions of cylinder sets, we have that $\mu(A' \cap T^{-n}B')$ is equal to $\mu(A')\mu(B')$ for large n , it is easy to deduce that $\mu(A \cap T^{-n}B) \rightarrow \mu(A)\mu(B)$ as required. Since the doubling map is measure-theoretically isomorphic to a one-sided Bernoulli shift, it follows that the doubling map is also strong mixing.

Similarly, if a Markov Chain is irreducible (i.e., for any states i and j , there exists an $n \geq 0$ such that $P_{ij}^n > 0$) and aperiodic (there is a state i such that $\gcd\{n; P_{ii}^n > 0\} = 1$), then given any pair of cylinder sets A' and B' , it follows from standard theorems of Markov chains $\mu(A' \cap T^{-n}B') \rightarrow \mu(A')\mu(B')$. The same argument as above then shows that an aperiodic irreducible Markov Chain is strong mixing. On the other hand, if a Markov chain is periodic ($d = \gcd\{n : P_{ii}^n > 0\} > 0$), then letting $A = B = \{x : x_0 = i\}$, it follows that $\mu(A \cap T^{-n}B) = 0$ whenever $d \nmid n$. Thus, T^d is not ergodic, so that T is not weak mixing. Note that a mixing Markov Chain is actually measure-theoretically isomorphic to a Bernoulli shift (Friedman and Ornstein 1970; Adler et al. 1972; Keane and Smorodinsky 1979b).

So far in this chapter, all examples of explicit mixing transformations have been isomorphic to Bernoulli transformations. Ornstein's construction of rank-one mixing using random spacers produced the first known examples of zero-entropy mixing (Ornstein 1972). Since these examples, tools have been developed to establish explicit rank-one mixing transformations (e.g., staircase transformations) (Adams 1998; Creutz and Silva 2010). Amazingly, it has already been established by Danilenko that any discrete countable infinite amenable group admits a zero-entropy mixing action (Danilenko 2016).

Both weak and strong mixing have formulations in terms of functions:

1. T is weak mixing if and only if for every $f, g \in L^2$ one has

$$\frac{1}{N} \sum_{n=0}^{N-1} |\langle f, g \circ T^n \rangle - \langle f, 1 \rangle \langle 1, g \rangle| \rightarrow 0 \text{ as } N \rightarrow \infty.$$

2. T is strong mixing if and only if for every $f, g \in L^2$, one has

$$\langle f, g \circ T^N \rangle \rightarrow \langle f, 1 \rangle \langle 1, g \rangle \text{ as } N \rightarrow \infty.$$

Using this, one can see that both mixing conditions are spectral properties.

Corollary 10 *Weak and strong mixing are spectral properties.*

This is an opportune time to highlight a couple major results that establish mixing properties for interval exchange transformations. In Katok (1980a), Katok shows that no interval exchange transformation is strong mixing. On the other hand, if the underlying permutation π is irreducible and not a rotation,² then the interval exchange transformation is weak mixing for almost all divisions of the interval (with respect to Lebesgue measure on the $k - 1$ -dimensional simplex). The case $k = 3$ was proved by Katok and Stepin (1967) and the general case including $k > 3$ was established by Avila and Forni (2007). (This research is cited as a primary contributor to Avila's 2014 Fields Medal (<https://www.mathunion.org/imu-awards/fields-medal/fields-medal-2014/fields-medallists-2014-awardees-brief-citations>).)

While on the face of it the formulation of weak mixing is considerably less natural than that of strong mixing, the notion of weak mixing turns out to be extremely natural from a spectral point of view. Given a measure-preserving transformation T , let U_T be the Koopman operator described above. Since this operator is an isometry, any eigenvalue must lie on the unit circle. The constant function 1 is always an eigenfunction with eigenvalue 1. Note that if $z = x + iy$ is in the complex

Lemma 9 *Let T be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$.*

² A permutation π on $\{1, 2, \dots, k\}$ is a rotation if $\pi(i) = i + 1 \pmod k$.

plane, then its complex conjugate $\bar{z} = x - iy$. If T is ergodic and g and h are eigenfunctions of U_T with eigenvalue λ , then $g\bar{h}$ is an eigenfunction with eigenvalue 1, hence invariant, so that $g = Kh$ for some constant K . Notice that for ergodic transformations, up to rescaling, there is at most one eigenfunction with any given eigenvalue.

If U_T has a nonconstant eigenfunction f , then one has $|\langle U_T^n f, f \rangle| = \|f\|^2$ for each n , whereas by Cauchy-Schwartz, $|\langle f, 1 \rangle|^2 < \|f\|^2$. It follows that $|\langle U_T^n f, f \rangle - \langle f, 1 \rangle \langle 1, f \rangle| \geq c$ for some positive constant c , so that using Lemma 9, T is not weak mixing.

The converse can be shown using the spectral theorem. For a detailed proof, see section 2.6 of Petersen (1983). Also, another chapter ► “Spectral Theory of Dynamical Systems” (Lemanczyk and Kanigowski 2022) is devoted to a study of spectral theory and its connections to ergodic theory. We encourage the interested reader to refer to this chapter for many of the detailed definitions and results on spectral theory.

Theorem 11 *The measure-preserving transformation T is weak mixing if and only if U_T has no nonconstant eigenfunctions.*

Of course this also shows that weak mixing is a spectral property. Equivalently, this says that the transformation T is weak mixing if and only if apart from the constant eigenfunction, the operator U_T has only continuous spectrum (i.e., the operator has no other eigenfunctions).

Using this theory, one can establish the following:

Theorem 12

1. T is weak mixing if and only if $T \times T$ is ergodic;
2. If T and S are ergodic, then $T \times S$ is ergodic if and only if U_S and U_T have no common eigenvalues other than 1.

For a measure-preserving transformation T , let K be the subspace of L^2 spanned by the eigenfunctions of U_T . It is a remarkable fact that K may be identified as $L^2(X, \mathcal{B}, \mu)$ where $\mathcal{B}'$ is a sub- σ -algebra of $\mathcal{B}$. The space K is called the *Kronecker factor* of T . The terminology comes

from the fact that any sub- σ -algebra $\mathcal{F}$ of $\mathcal{B}$ gives rise to a factor mapping $\pi : (X, \mathcal{B}, \mu) \rightarrow (X, \mathcal{F}, \mu)$ with $\pi(x) = x$. By construction $L^2(X, \mathcal{B}', \mu)$ is the closed linear span of the eigenfunctions of T considered as a measure-preserving transformation of $(X, \mathcal{B}', \mu)$. By the Discrete Spectrum Theorem of Halmos and von Neumann (1942), T acting on $(X, \mathcal{B}', \mu)$ is measure-theoretically isomorphic to a rotation on a compact group. This allows one to split $L^2(X, \mathcal{B}, \mu)$ as $L^2(X, \mathcal{B}, \mu) \oplus L_c^2(X, \mathcal{B}, \mu)$, where, as mentioned above the first part is the discrete spectrum part, spanned by eigenfunctions, and the second part is the continuous spectrum part, consisting of functions whose spectral measure is continuous. Since L^2 is split into a discrete part and a continuous part, it is natural to ask whether the underlying transformation T can be split up in some way into a weak mixing part and a discrete spectrum (compact group rotation) part, somewhat analogously to the ergodic decomposition. Unfortunately, there is no such decomposition available. However, for some applications, for example, to multiple recurrence (starting with the work of Furstenberg (1977, 1981)), the decomposition of L^2 (possibly into more complicated parts) plays a crucial role (see the chapters “Ergodic Theory: Recurrence” and “Ergodic Theory: Interactions with Combinatorics and Number Theory”).

For noninvertible measure-preserving transformations, the transformation is weak or strong mixing if and only if its natural extension has that property.

The understanding of weak mixing in terms of the discrete part of the spectrum of the operator also extends to total ergodicity. T^n is ergodic if and only if T has no eigenvalues of the form $e^{2\pi i p/q}$ other than 1. From this, it follows that an ergodic measure-preserving transformation T is totally ergodic if and only if it has no *rational spectrum* (i.e., no eigenvalues of the form $e^{2\pi i p/q}$ other than the simple eigenvalue 1).

An intermediate mixing condition between strong- and weak mixing is that a measure-preserving transformation is *mild-mixing* if whenever $f \circ T^{n_i} \rightarrow f$ for an L^2 function f and a sequence $n_i \rightarrow \infty$, then f is a.e. constant. Clearly mild-mixing is a spectral property. If a

transformation has an eigenfunction f , then it is straightforward to find a sequence n_i such that $f \circ T^{n_i} \rightarrow f$, thus, it follows that mild-mixing implies weak mixing. To see that strong mixing implies mild-mixing, suppose that T is strong mixing and that $f \circ T^{n_i} \rightarrow f$. Then $\int f \circ T^{n_i} f \rightarrow \|f\|^2$. On the other hand, the strong mixing property implies that $\int f \circ T^{n_i} \bar{f} \rightarrow |\langle f, 1 \rangle|^2$. The equality of these implies that f is a.e. constant. Mild-mixing has a useful reformulation in terms of ergodicity of general (not necessarily probability) measure-preserving transformations: A transformation T is mild-mixing if and only if for every conservative ergodic measure-preserving transformation S , $T \times S$ is ergodic. See Furstenberg and Weiss' article (1978) for further information on mild-mixing.

If there exists a sequence $n_i \rightarrow \infty$ such that for every $f \in L^2$, $f \circ T^{n_i} \rightarrow f$, then T is said to be rigid. The sequence n_i is called a rigidity sequence for T . Mild mixing transformations are those transformations with no rigid factor. There has been much new research on the nature of rigidity sequences, spurred largely by (Bergelson et al. 2014). See (Eisner and Grivaux 2011; Aaronson et al. 2014; Adams 2015; Fayad and Kanigowski 2015; Grivaux 2013; Le 2017; Bayless and Yancey 2015; Grivaux and Roginskaya 2013; Robertson 2019) for a selection of recent results on rigidity and nonrecurrent sequences. If there exists $\alpha > 0$ such that for every measurable set A ,

$$\liminf_{n \rightarrow \infty} \mu(A \cap T^{-n}A) \geq \alpha \mu(A),$$

then T is said to be partially rigid, or more specifically α -rigid. Strong mixing transformations are not partially rigid for any $\alpha > 0$. While mild mixing transformations are not rigid, they can be partially rigid, and in particular, both Chacon transformations (Chacon-2 and Chacon-3) are partially rigid.

In the case where the space X is endowed with a topology, a transformation T is topologically mixing if given two nonempty open sets U and V , there exists $N \in \mathbb{N}$ such that for all $n > N$, the set $V \cap T^{-n}U$ is nonempty. Topological mixing does not imply ergodicity (Muehlegger et al. 1997). If every open set has positive measure,

then strong mixing implies topological mixing. Mild mixing does not imply topological mixing. There is a measure theoretic property which is a close counterpart to topological mixing. The property is commonly referred to as light mixing and was first introduced by Walters as intermixing (Walters 1972). A transformation T is lightly mixing if given any two sets A and B of positive measure,

$$\liminf_{n \rightarrow \infty} \mu(A \cap T^{-n}B) > 0.$$

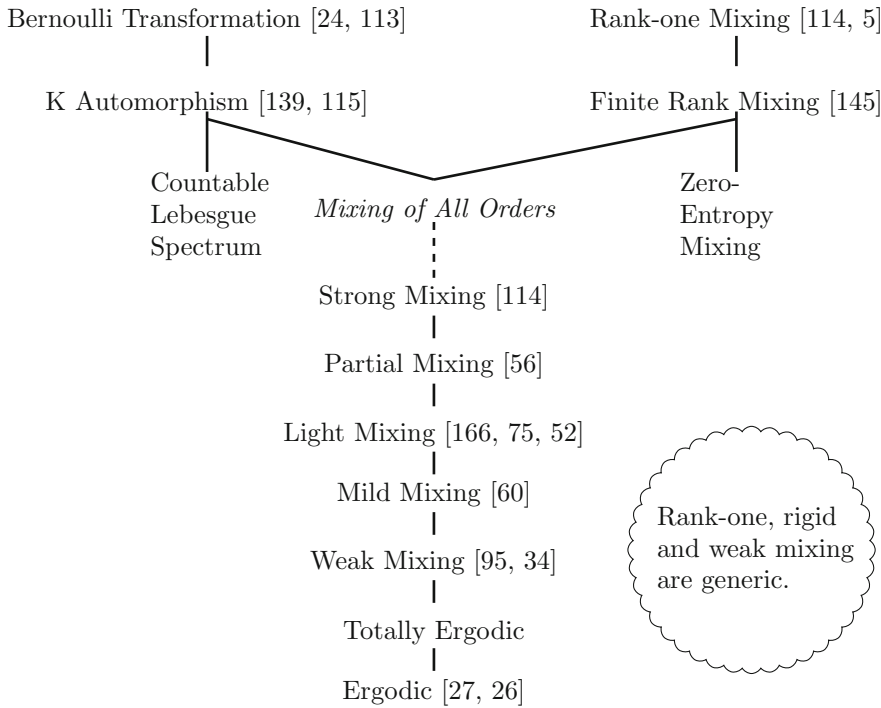
If T is lightly mixing, then T is mildly mixing (since $\liminf_{n \rightarrow \infty} \mu(A \cap T^{-n}A^c) = 0$ for A in a rigid factor). Also, there exist transformations which are lightly mixing, but are not strong mixing. In particular, the transformation first published by Chacon as a weak mixing, nonstrong mixing transformation is actually lightly mixing (Chacon 1969; Friedman and King 1991).

If there exists $\alpha > 0$ such that for all measurable sets A and B ,

$$\liminf_{n \rightarrow \infty} \mu(A \cap T^{-n}B) \geq \alpha \mu(A) \mu(B),$$

then T is partially mixing, or more specifically α -mixing. It is not difficult to show that $\alpha \leq 1$ for T measure-preserving. There are transformations which are partially mixing, but are not strongly mixing. Also, Chacon's transformation from (Friedman and King 1991) is lightly mixing, but not partially mixing. In Friedman (1989), Friedman constructs explicit transformations with an optimal combination of mixing and rigidity. In particular, for each α such that $0 < \alpha < 1$, a transformation is constructed that is simultaneously α -rigid and $(1 - \alpha)$ -mixing.

See Fig. 2 for a hierarchy of mixing properties. The stronger more restrictive properties start at the top and become weaker as one descends down the tree. Note that there are known examples distinguishing these properties, except in the case of strong mixing and mixing of all orders. The dashed line between these properties signifies that it is an open question whether strong mixing (i.e., 2-mixing) implies mixing of all orders. Many of the branches are labeled with references to



Ergodicity and Mixing Properties, Fig. 2 Mixing hierarchy

articles where these properties were defined or important examples were presented.

The strongest spectral property considered is that of having countable Lebesgue spectrum. A detailed discussion of spectral theory is not included here, although this special case that can be described simply. Specifically, let T be an invertible measure-preserving transformation. Then T has *countable Lebesgue spectrum* if there is a sequence of functions $f_1, f_2, \dots$ such that $\{1\} \cup \{U_T^n f_j : n \in \mathbb{Z}, j \in \mathbb{N}\}$ forms an orthonormal basis for $L^2(X)$.

To see that this property is stronger than strong mixing, simply observe that it implies that $\langle U_T^t U_T^n f_j, U_T^m f_k \rangle \rightarrow 0$ as $t \rightarrow \infty$. Then by approximating f and g by their expansions with respect to a finite part of the basis, it can be deduced that $\langle U_T^n f, g \rangle \rightarrow \langle f, 1 \rangle \langle 1, g \rangle$ as required. Since already strong mixing is atypical from the topological point of view, it follows that countable Lebesgue spectrum has to be atypical. In fact, Yuzvinskii (1967) showed that the typical

invertible measure-preserving transformation has simple singular spectrum.

The property of countable Lebesgue spectrum is by definition a spectral property. Since it completely describes the transformation up to spectral isomorphism, there can be no stronger spectral properties. The remaining properties that are examined are invariant under measure-theoretic isomorphisms only.

An invertible measure-preserving transformation T of $(X, \mathcal{B}, \mu)$ is said to be K (for *Kolmogorov*) if there is a sub- σ -algebra $\mathcal{F}$ of $\mathcal{B}$ such that

1. $\bigcap_{n=1}^{\infty} T^{-n} \mathcal{F}$ is the trivial σ -algebra up to sets of measure 0 (i.e., the intersection consists only of null sets and sets of full measure).
2. $\bigvee_{n=1}^{\infty} T^n \mathcal{F} = \mathcal{B}$ (i.e., the smallest σ -algebra containing $T^n \mathcal{F}$ for all $n > 0$ is $\mathcal{B}$).

The K property has a useful reformulation in terms of entropy as follows: T is K if and only if for every nontrivial partition $\mathcal{P}$ of X , the entropy of T with respect to the partition $\mathcal{P}$ is positive:

T has *completely positive entropy*. See the chapter on Entropy in Ergodic Theory for the relevant definitions. The equivalence of the K property and completely positive entropy was shown by Rokhlin and Sinai (1961). For a general transformation T , one can consider the collection of all subsets B of X such that with respect to the partition $\mathcal{P}_B = \{B, B^c\}$, $h(\mathcal{P}_B) = 0$. One can show that this is a σ -algebra. This σ -algebra is known as the *Pinsker σ -algebra*. The above reformulation allows us to say that a transformation is K if and only if it has a trivial Pinsker σ -algebra.

The K property implies countable Lebesgue spectrum (see Parry's (1981) book for a proof). To see that K is not implied by countable Lebesgue spectrum, it can be pointed out that certain transformations derived from Gaussian systems (see, for example, the paper of Newton and Parry (1966)) have countable Lebesgue spectrum but zero-entropy.

The fact that (two-sided) Bernoulli shifts have the K property follows from Kolmogorov's 0–1 law by taking $\mathcal{F} = \bigvee_{n=0}^{\infty} T^{-n} \mathcal{P}$, where $\mathcal{P}$ is the partition into cylinder sets (see Williams's (1991) book for details of the 0–1 law).

Although the K property is explicitly an invertible property, it has a noninvertible counterpart, namely, exactness. A transformation T of $(X, \mathcal{B}, \mu)$ is *exact* if $\bigcap_{n=0}^{\infty} T^{-n} \mathcal{B}$ consists entirely of null sets and sets of measure 1. It is not hard to see that a noninvertible transformation is exact if and only if its natural extension is K.

The final and strongest property in our list is that of being measure-theoretically isomorphic to a Bernoulli shift. If T is measure-theoretically isomorphic to a Bernoulli shift, it can be said that T has the *Bernoulli property*. While in principle this could apply to both invertible and noninvertible transformations, in practice the definition applies to a large class of invertible transformations, but occurs comparatively seldom for noninvertible transformations. For this reason, we will restrict ourselves to a discussion of the Bernoulli property for invertible transformations (see however work of Hoffman and Rudolph (2002) and Hecklen and Hoffman (2002) for work on the one-sided Bernoulli property).

In the case of invertible Bernoulli shifts, Ornstein (1970, 1974) developed in the early 1970s a powerful isomorphism theory, showing that two Bernoulli shifts are measure-theoretically isomorphic if and only if they have the same entropy. Entropy had already been identified as an invariant by Kolmogorov and Sinai (Kolmogorov 1958; Sinai 1959), so this established that it was a complete invariant for Bernoulli shifts. Keane and Smorodinsky (1979a) gave a proof which showed that two Bernoulli shifts of the same entropy are isomorphic using a conjugating map that is continuous almost everywhere. With other authors, this theory was extended to show that the property of being isomorphic to a Bernoulli shift applied to a surprisingly large class of measure-preserving transformations (e.g., geodesic flows on manifolds of constant negative curvature (Ornstein and Weiss 1973), aperiodic irreducible Markov chains (Friedman and Ornstein 1970), toral automorphisms (Katznelson 1971), and more generally many Gibbs measures for hyperbolic dynamical systems (see the book of Bowen (1975))).

Initially, it was conjectured that the properties of being K and Bernoulli were the same, but since then a number of measure-preserving transformations that are K but not Bernoulli have been identified. The earliest was due to Ornstein (1973a). Ornstein and Shields (1973) then provided an uncountable family of nonisomorphic K automorphisms. Katok (1980b) gave an example of a smooth diffeomorphism that is K but not Bernoulli, and Kalikow (1982) gave a very natural probabilistic example of a transformation that has this property (the T, T^{-1} process).

While in systems that one regularly encounters there is a correlation between positive entropy and the stronger mixing properties discussed, these properties are logically independent; for example, taking the product of a Bernoulli shift and the identity transformation gives a positive entropy transformation that fails to be ergodic; also, besides rank-one mixing examples, there are zero-entropy Gaussian systems which are strong mixing and have countable Lebesgue spectrum.

In many of the mixing criteria discussed above, a pair of sets A and B is considered and one asks for asymptotic independence of A and B (so that for large n , A and $T^{-n} B$ become independent).

It is natural to ask, given a finite collection of sets $A_0, A_1, \dots, A_k$, under what conditions $\mu(A_0 \cap T^{-n_1}A_1 \cap \dots \cap T^{-n_k}A_k)$ converges to $\prod_{j=0}^k \mu(A_j)$.

A measure-preserving transformation T is said to be *mixing of order $k+1$* if for all measurable sets $A_0, \dots, A_k$,

$$\lim_{n_1 \rightarrow \infty, n_{j+1} - n_j \rightarrow \infty} \mu(A_0 \cap T^{-n_1}A_1 \cap \dots \cap T^{-n_k}A_k) = \prod_{j=0}^k \mu(A_j).$$

The transformation T is said to be *mixing of all orders*, if T is mixing of order $k+1$ for all natural numbers k . An outstanding open question asked by Rokhlin (1949) appearing already in Halmos's (1956) book is to determine whether mixing (i.e., mixing of order 2) implies mixing of all orders. Kalikow (1984) showed that mixing implies mixing of all orders for rank-one transformations (existence of rank-one mixing transformations having been previously established by Ornstein (1972)). Later Ryzhikov (1993) used joining methods to establish the result for transformations with finite rank, and Host (1991) also used joining methods to establish the result for measure-preserving transformations with singular spectrum, but the general question remains open.

It is not hard to show using martingale arguments that K automorphisms and hence all Bernoulli measure-preserving transformations are mixing of all orders.

In Furstenberg (1981), Furstenberg defined weak mixing of all orders as a natural analogue to the stronger condition of mixing of all orders. A measure-preserving transformation T is *weak mixing of all orders*, if given measurable sets $A_0, \dots, A_k$, there is a subsequence J of the integers of density 0 such that

$$\lim_{n \rightarrow \infty, n \notin J} \mu(A_0 \cap T^{-n}A_1 \cap \dots \cap T^{-kn}A_k) = \prod_{i=0}^k \mu(A_i).$$

In Furstenberg (1977), Furstenberg proved that weak mixing implies weak mixing of all orders.

Bergelson (1987) generalized this by showing that weak mixing implies a polynomial version of weak mixing of all orders:

$$\lim_{n \rightarrow \infty, n \notin J} \mu(A_0 \cap T^{-p_1(n)}A_1 \cap \dots \cap T^{-p_k(n)}A_k) = \prod_{i=0}^k \mu(A_i)$$

whenever $p_1(n), \dots, p_k(n)$ are nonconstant integer-valued polynomials such that $p_i(n) - p_j(n)$ is unbounded for $i \neq j$. The proofs in Furstenberg (1977) and Bergelson (1987) make use of a Hilbert space version of the van der Corput inequality of analytic number theory. Moreover, this method is instrumental in Furstenberg's ergodic proof (Furstenberg 1977) of Szemerédi's theorem on the existence of arbitrarily long arithmetic progressions in a subset of the integers of positive density (see the chapter "Ergodic Theory: Interactions with Combinatorics and Number Theory" for more information about this direction of study).

The conclusions that one draws here are much weaker than the requirement for mixing of all orders. For mixing of all orders, it was required that provided the gaps between $0, n_1, \dots, n_k$ diverge to infinity, one achieves asymptotic independence, whereas for these weak mixing results, the gaps are increasing along prescribed sequences with regular growth properties.

It is interesting to note that the analogous question of whether mixing implies mixing of all orders is known to fail in higher-dimensional actions. Here, rather than a $\mathbb{Z}$ action, in which there is a single measure-preserving transformation (so that the integer n acts on a point $x \in X$ by mapping it to $T^n x$), one takes a $\mathbb{Z}^d$ action. For such an action, one has d commuting transformations $T_1, \dots, T_d$ and a vector $(n_1, \dots, n_d)$ acts on a point x by sending it to $T_1^{n_1} \dots T_d^{n_d} x$. Ledrappier (1978) studied the following two-dimensional action. Let $X = \{x \in \{0, 1\}^{\mathbb{Z}^2} : x_v + x_{v+e_1} + x_{v+e_2} = 0 \pmod{2}\}$ and let $T_i(x)_v = x_{v+e_i}$. Since X is a compact Abelian group, it has a natural measure μ invariant under the group operations (the Haar measure). It is not hard to show that this

system is mixing (i.e., given any measurable sets A and B , $\mu(A \cap T_1^{-n_1} T_2^{-n_2} B) \rightarrow \mu(A)\mu(B)$ as $\|(n_1, n_2)\| \rightarrow \infty$). Ledrappier showed that the system fails to be 3-mixing. Subsequently Masser (2004) established necessary and sufficient conditions for similar higher-dimensional algebraic actions to be mixing of order k but not order $k + 1$ for any given k .

Hyperbolicity and Decay of Correlations

One class of systems in which the stronger mixing properties are often found is the class of smooth systems possessing uniform hyperbolicity (i.e., the tangent space to the manifold at each point splits into stable and unstable subspaces $E^s(x)$ and $E^u(x)$ such that the $\left\|DT\right|_{E^s(x)}\right\| \leq a < 1$ for all x and $\left\|DT^{-1}\right|_{E^u(x)}\right\| \leq a$ and $DT(E^s(x)) = E^s(T(x))$ and $DT(E^u(x)) = E^u(T(x))$). In some cases, similar conclusions are found in systems possessing nonuniform hyperbolicity. See Katok and Hasselblatt's (1995) book for an overview of hyperbolic dynamical systems, as well as the chapter in this volume on ► [“Smooth Ergodic Theory.”](#)

In the simple case of expanding piecewise continuous maps of the interval (i.e., maps for which the absolute value of the derivative is uniformly bounded below by a constant greater than 1), it is known that if they are totally ergodic and topologically transitive (i.e., the forward images of any interval cover the entire interval), then provided that the map has sufficient smoothness (e.g., the map is C^1 and the derivative satisfies a certain additional summability condition), the map has a unique absolutely continuous invariant measure which is exact and whose natural extension is Bernoulli (see the paper of Góra (1994) for results of this type proved under some of the mildest hypotheses). These results were originally established for maps that were twice continuously differentiable, and the hypotheses were progressively weakened, approaching, but never meeting, C^1 . Subsequent work of Quas (1996b, a) provided

examples of C^1 expanding maps of the interval for which Lebesgue measure was invariant, but respectively not ergodic and not weak mixing. Some of the key tools in controlling mixing in one-dimensional expanding maps that are absent in the C^1 case are bounded distortion estimates. Here, there is a constant $1 \leq C < \infty$ such that given any interval I on which some power T^n of T acts injectively and any subinterval J of I , one has $1/C \leq (|T^n J| / |T^n I|) / (|J| / |I|) \leq C$. An early place in which bounded distortion estimates appear is the work of Rényi (1957).

One important class of results for expanding maps establishes an exponential decay of correlations. Here, one starts with a pair of smooth functions f and g and one estimates $\int f \cdot g \circ T^n d\mu - \int f d\mu \int g d\mu$, where μ is an absolutely continuous invariant measure. If μ is mixing, it is expected that this will converge to 0. In fact though, in good cases this converges to 0 at an exponential rate for each pair of functions f and g belonging to a sufficiently smooth class. In this case, the measure-preserving transformation T is said to have exponential decay of correlations. See Liverani's (2004) article for an introduction to a method of establishing this based on cones. Exponential decay of correlations implies in particular that the natural extension is Bernoulli.

Hu (2004) has studied the situation of maps of the interval for which the derivative is bigger than 1 everywhere except at a fixed point, where the local behavior is of the form $x \mapsto x + x^{1+\alpha}$ for $0 < \alpha < 1$. In this case, rather than exhibiting exponential decay of correlations, the map has polynomial decay of correlations with a rate depending on α .

In Young's (1995) survey, a variety of techniques are outlined for understanding the strong ergodic properties of nonuniformly hyperbolic diffeomorphisms. In her article (Young 1998), methods are introduced for studying many classes of nonuniformly hyperbolic systems by looking at suitably high powers of the map, for which the power has strong hyperbolic behavior. The article shows how to understand the ergodic behavior of these systems. These methods are applied (for

example) to billiards, one-dimensional quadratic maps, and Hénon maps.

Representations, Realizations, and Genericity

There are multiple representations of ergodic invertible measure-preserving transformations. One of the most common representations is as a shift transformation on doubly infinite sequences of characters from a finite alphabet. This can be made into a compact metric space by setting the distance between two infinite sequences $x = (x_i)$ and $y = (y_i)$ as

$$d(x, y) = |x_0 - y_0| + \sum_{i=1}^{\infty} 2^{-i} (|x_i - y_i| + |x_{-i} - y_{-i}|). \quad (3)$$

If T has zero-entropy, then there is a two set partition $P = \{p_0, p_1\}$ such that $\bigvee_{i=0}^{\infty} T^{-i}P$ generates the sigma algebra. Pick a point x that is typical for T , and let $x_i = 0$, if $T_x^i \in p_0$, otherwise set $x_i = 1$. Let $\bar{X}$ be the closure of all left and right shifts of the sequence x . The left shift operator on the sequences in $\bar{X}$ is isomorphic to T . This procedure can be extended to realize any ergodic invertible measure-preserving transformation with finite entropy. They can all be realized by a shift on a compact metric space (Jewett 1970; Krieger et al. 1972). The shift will be continuous in the topology induced by the distance d defined in 3. Using an inverse limit of refining partitions, it is possible to construct a homeomorphism realization of T . Also, T can be constructed such that it is strictly ergodic which means there is only one invariant ergodic measure, and every point is typical.

Later, it was shown by Lehrer (1987) that it is possible to obtain an isomorphic transformation that is strictly ergodic and topologically mixing. This is further validation that topological mixing is not a measure theoretic isomorphic invariant.

Other representations of ergodic invertible measure-preserving transformations are as infinite

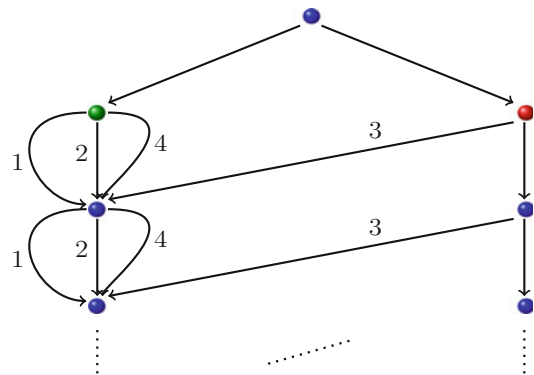
interval exchange transformations. This is an extension of finite interval exchange transformations except the domain and range are partitioned into a countable number of intervals. It was shown in Arnoux and Weiss (1985) that every ergodic invertible measure-preserving transformation on a Lebesgue probability space is isomorphic to an infinite interval exchange transformation.

A similar but itself useful technique for constructing a transformation is through cutting and stacking. It was also shown in Arnoux and Weiss (1985) that every ergodic invertible measure-preserving transformation on a Lebesgue probability space has an isomorphic version which is obtained by cutting and stacking. It may be that the supremum of the number of columns goes to infinity. In this case, if there is not an isomorphic version with a bounded number of columns, the transformation has infinite rank.

Finally, another method for obtaining an ergodic invertible measure-preserving transformation is via an adic transformation on a Bratteli-Vershik diagram (Vershik 1981; Vershik and Livshits 1992). It is similar to the cutting and stacking procedure, although adic transformations encode points as infinite paths in a directed acyclic graph. Figure 3 displays an adic representation of the Chacon-3 transformation.

Topological and Set-Theoretic Properties of Classes of Transformations

It is of interest to understand the behavior of a “typical” measure-preserving transformation.



Ergodicity and Mixing Properties, Fig. 3 Chacon-3 Adic

There are a number of Baire category results addressing this. In order to state them, one needs a set of measure-preserving transformations and a topology on them. As mentioned earlier, it is effectively no restriction to assume that a transformation is a Lebesgue-measurable map on the unit interval preserving Lebesgue measure. The classical category results are then on the collection of invertible Lebesgue-measure-preserving transformations of the unit interval. One topology on these is the “weak” topology, where a subbase is given by sets of the form $N(T, A, \epsilon) = \{S; \lambda(S(A) \Delta T(A)) < \epsilon\}$. With respect to this topology, Halmos (1944) showed that a residual set (i.e., a dense G_δ set) of invertible measure-preserving transformations is weak mixing (see also work of Alpern (1976)), while Rokhlin (1948) showed that the set of strong mixing transformations is meagre (i.e., a nowhere dense F_σ set), allowing one to conclude that with respect to this topology, a typical transformation is weak but not strong mixing.

In Tikhonov (2007), Tikhonov introduces a natural metric (i.e., leash topology) which makes the set of mixing transformations into a complete separable metric space. It is shown that in this metric, a generic mixing transformation is mixing of all orders and has simple singular spectrum and all its powers are disjoint. See the chapter ▶ “Spectral Theory of Dynamical Systems” (Lemanczyk and Kanigowski 2022) for further details on simple singular spectrum, as well as the notion of disjointness. While it is shown that the conjugacy class of a generic mixing transformation is dense, not until Bashtanov (2013), is it shown that the conjugacy class of every mixing transformation is dense in the leash topology. Also, in Bashtanov (2013), it is shown that rank-one is generic in the leash topology.

Viewing classes of ergodic measure-preserving transformations from a descriptive set theoretic viewpoint (Becker and Kechris 1996; Foreman 2000) has shed new light on the challenges characterizing broad classes of transformations. In particular, in Foreman and Weiss (2011), it is shown that the isomorphism relation is a complete analytic set and in particular, not Borel. This gives an obstacle to concrete (countably) verifiable

invariants for determining when two transformations are isomorphic. However, it is also shown that there are topologically generic classes of transformations (i.e., rank-one) that are Borel. At this time, it is an open question to establish an easily verifiable method for determining when two rank-one transformations are isomorphic. Nevertheless, there has been recent progress toward characterizing rank-one transformations; see Gao and Hill (2014, b); Adams et al. (2017); and Foreman et al. (2019).

Future Directions

Problem 1 (Mixing of all orders). Does mixing imply mixing of all orders? Can the results of Kalikow, Ryzhikov, and Host be extended to larger classes of measure-preserving transformations? Thouvenot observed that it is sufficient to establish the result for measure-preserving transformations of entropy 0. This observation (whose proof is based on the Pinsker σ -algebra) was stated in Kalikow’s (1984) paper and is reproduced as Proposition 3.2 in recent work of de la Rue (2006) on the mixing of all orders problem.

Problem 2 (Multiple weak mixing). As mentioned above, Bergelson (1987) showed that if T is a weak mixing transformation, then there is a subset J of the integers of density 0 such that

$$\lim_{n \rightarrow \infty, n \notin J} \mu \left(A_0 \cap T^{-p_1(n)} A_1 \cap \dots \cap T^{-p_k(n)} A_k \right) = \prod_{i=0}^k \mu(A_i)$$

whenever $p_1(n), \dots, p_k(n)$ are nonconstant integer-valued polynomials such that $p_i(n) - p_j(n)$ is unbounded for $i \neq j$. It is natural to ask what is the most general class of times that can replace the sequences $(p_1(n)), \dots, (p_k(n))$. In unpublished notes, Bergelson and Håland considered as times the values taken by a family of integer-valued generalized polynomials (those functions of an integer variable that can be obtained by the

operations of addition, multiplication, addition of or multiplication by a real constant and taking integer parts (e.g., $g(n) = \lfloor \sqrt{2} \lfloor \pi n \rfloor + \lfloor \sqrt{3n} \rfloor$). They conjectured necessary and sufficient conditions for the analogue of Bergelson's weak mixing polynomial ergodic theorem to hold and proved the conjecture in certain cases.

In a recent paper of McCutcheon and Quas (2007), the analogous question was addressed in the case where T is a mild-mixing transformation.

Problem 3 (Pascal adic transformation). Vershik (1974, 1981) introduced a family of transformations known as the adic transformations. The underlying spaces for these transformations are certain spaces of paths on infinite graphs and the transformations act by taking a path to its lexicographic neighbor. Among the adic transformations, the so-called Pascal adic transformation (so-called because the underlying graph resembles Pascal's triangle) has been singled out for attention in work of Petersen and others (Petersen and Schmidt 1997; Méla and Petersen 2005; Berend and Kolesnik 2001; Adams and Petersen 1998). In particular, it is unresolved whether this transformation is weak mixing with respect to any of its ergodic measures. Weak mixing has been shown to follow from a number-theoretic condition on the binomial coefficients (Berend and Kolesnik 2001; Adams and Petersen 1998).

Problem 4 (Weak Pinsker Conjecture - SOLVED). Pinsker (1960) conjectured that in a measure-preserving transformation with positive entropy, one could express the transformation as a product of a Bernoulli shift with a system with zero entropy. This conjecture (now known as the *Strong Pinsker Conjecture*) was shown to be false by Ornstein (1973b, 1973c). Shields and Thouvenot (1975) showed that the collection of transformations that can be written as a product of a zero-entropy transformation with a Bernoulli shift is closed in the so-called $\bar{d}$ -metric that lies at the heart of Ornstein's theory.

It is, however, the case that if $T: X \rightarrow X$ has entropy $h > 0$, then for all $h' \leq h$, T has a factor S with entropy h' (this was originally proved by Sinai (1964) and reproved using the Ornstein

machinery by Ornstein and Weiss (1975)). The *Weak Pinsker Conjecture* states that if a measure-preserving transformation T has entropy $h > 0$, then for all $\epsilon > 0$, T may be expressed as a product of a Bernoulli shift and a measure-preserving transformation with entropy less than ϵ .

This problem was recently solved in the affirmative by T. Austin (2018). New results on measure concentrations were developed to prove the weak Pinsker conjecture.

References

- Aaronson J (1997) An introduction to infinite ergodic theory. American Mathematical Society
- Aaronson J, Hosseini M, Lemanczyk M (2014) Ip-rigidity and eigenvalue groups. *Ergodic Theory Dynam Syst* 34(4):105–1076
- Abramov L (1959) The entropy of a derived automorphism. *Dokl Akad Nauk SSSR* 128:647–650
- Adams TM (1998) Smorodinsky's conjecture on rank-one mixing. *Proc Am Math Soc* 126(3):739–744
- Adams TM (2015) Tower multiplexing and slow weak mixing. *Colloq Math* 1(138):47–71
- Adams TM, Nobel AB (2010) Uniform convergence of vapnik-chervonenkis classes under ergodic sampling. *Ann Probab* 38(4):1345–1367
- Adams TM, Petersen K (1998) Binomial coefficient multiples of irrationals. *Monatsh Math* 125:269–278
- Adams TM, Ferenczi S, Petersen K (2017) Constructive symbolic presentations of rank one measure-preserving systems. *Colloq Math* 150:243–255
- Adler RL, Shields P, Smorodinsky M (1972) Irreducible markov shifts. *Ann Math Stat* 43(3):1027–1029
- Alpern S (1976) New proofs that weak mixing is generic. *Invent Math* 32:263–278
- Ambrose W, Kakutani S (1942) Structure and continuity of measurable flows. *Duke Math J* 9(1):25–42
- Arnoux DSOP, Weiss B (1985) Cutting and stacking, interval exchanges and geometric models. *Israel J Math* 50: 160–168
- Arora S, Ge R, Neyshabur B, Zhang Y (2018) Stronger generalization bounds for deep nets via a compression approach. In: Krause A, Dy J (eds) 35th international conference on machine learning, ICML 2018. International Machine Learning Society (IMLS), pp 390–418
- Arora S, Du S, Hu W, Li Z, Wang R (2019) Fine-grained analysis of optimization and generalization for over-parameterized two-layer neural networks. In: Chaudhuri K, Salakhutdinov R (eds) Proceedings of the 36th international conference on machine learning, volume 97 of proceedings of machine learning research. PMLR, Long Beach, pp 322–332
- Austin T (2010) On the norm convergence of non-conventional ergodic averages. *Ergodic Theory Dynam Syst* 30(2):321–338

- Austin T (2018) Measure concentration and the weak Pinsker property. *Publications mathématiques de l'IHÉS* 128(1):1–119
- Avila A, Forni G (2007) Weak-mixing for interval exchange transformations and translation flows. *Ann Math* 165:637–664
- Bashtanov AI (2013) Generic mixing transformations are rank 1. *Math Notes* 93:209–216
- Bayless RL, Yancey KB (2015) Weakly mixing and rigid rank-one transformations preserving an infinite measure. *New York J Math* 21:615–636
- Becker H, Kechris AS (1996) The descriptive set theory of Polish group actions, volume 232. Cambridge University Press
- Berend MBD, Kolesnik G (2001) Irrational dilations of Pascal's triangle. *Mathematika* 48:159–168
- Bergelson V (1987) Weakly mixing per. *Ergodic Theory Dynam Syst* 7:337–349
- Bergelson V, Leibman A (2002) A nilpotent Roth theorem. *Inventiones mathematicae* 147(2):429–470
- Bergelson V, del Junco A, Lemańczyk M, Rosenblatt J (2014) Rigidity and non-recurrence along sequences. *Ergodic Theory Dynam Syst* 34(5):1464–1502
- Birkhoff GD (1931) Proof of the ergodic theorem. *Proc Nat Acad Sci* 17:656–660
- Birkhoff GD, Smith PA (1924) Structural analysis of surface transformations. *J Undergrad Math* 7:345–379
- Boltzmann L (1871) Einige allgemeine Sätze über Wärmegleichgewicht. *Wiener Berichte* 63:679–711
- Boltzmann L (1909) *Wissenschaftliche Abhandlungen*. Akad. Wissenschaften Berlin
- Bowen R (1975) *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*. Springer
- Bozgan F, Sanchez A, Silva CE, Stevens D, Wang J (2015) Subsequence bounded rational ergodicity of rank-one transformations. *Dyn Syst* 30(1):70–84
- Bradley RC (2005) Basic properties of strong mixing conditions. A survey and some open questions. *Probab Surv* 2:107–144
- Chacon RV (1966a) Change of velocity in flows. *J Math Mechanics* 16(5):417–431
- Chacon RV (1966b) Transformations having continuous spectrum. *J Math Mechanics* 16(5):399–415
- Chacon RV (1969) Weakly mixing transformations which are not strongly mixing. *Proc Am Math Soc* 22:559–562
- Choquet G (1956a) Existence des représentations intégrales au moyen des points extrémaux dans les cônes convexes. *C R Acad Sci Paris* 243:699–702
- Choquet G (1956b) Unité des représentations intégrales au moyen de points extrémaux dans les cônes convexes réticulés. *C. R. Acad Sci Paris* 243:555–557
- Conze J-P, Lesigne E (1984) Théorèmes ergodiques pour des mesures diagonales. *Bull de la Soc Math de France* 112:143–175
- Creutz D, Silva CE (2010) Mixing on rank-one transformations. *Stud Math* 199(1):43–72
- Danilenko AI (2016) Mixing actions of zero entropy for countable amenable groups. In: *Colloquium Mathematicum*, volume 145. Instytut Matematyczny Polskiej Akademii Nauk, pp 179–186
- de la Rue T (2006) 2-fold and 3-fold mixing: why 3-dot-type counterexamples are impossible in one dimension. *Bull Braz Math Soc (NS)* 37(4):503–521
- Du SS, Zhai X, Poczos B, Singh A (2019) Gradient descent provably optimizes over-parameterized neural networks. In: *International conference on learning representations* 2019
- Ehrenfest P, Ehrenfest T (1911) *Begriffliche Grundlage der statistischen Auffassung in der Mechanik*. Number 4 in *Encycloplädie der mathematischen Wissenschaften*. Teubner
- Eisner T, Grivaux S (2011) Hilbertian Jamison sequences and rigid dynamical systems. *J Funct Anal* 261(7):2013–2052
- Fayad B, Kanigowski A (2015) Rigidity times for a weakly mixing dynamical system which are not rigidity times for any irrational rotation. *Ergodic Theory Dynam Syst* 35(8):2529–2534
- Feller W (1950) *An introduction to probability and its applications*. Wiley
- Ferenczi S (1997) Systems of finite rank. *Colloq Math* 73(1):35–65
- Fields medalists 2014 awardees with brief citations. <https://www.mathunion.org/imu-awards/fields-medal/fields-medal-2014/fields-medallists-2014-awardees-brief-citations>
- Foreman M (2000) A descriptive view of ergodic theory. *Descriptive Set Theory Dynam Syst*:87–171
- Foreman DJRM, Weiss B (2011) The conjugacy problem in ergodic theory. *Ann Math* 173(3):1529–1586
- Foreman M et al (2019) Rank-one transformations, odometers, and finite factors. *arXiv:1910.126J5*, pp 1–20
- Frantzikinakis N, Kra B (2005) Convergence of multiple ergodic averages for some commuting transformations. *Ergodic Theory Dynam Syst* 25(3):799–809
- Friedman NA (1970) *Introduction to ergodic theory*, Van Nostrand Reinhold Mathematical Studies, No. 29. Van Nostrand Reinhold Co, New York-Toronto, Ont.-London
- Friedman N (1989) Partial mixing, partial rigidity, and factors. *Contemp Math* 94:141–145
- Friedman N (1992) Replication and stacking in ergodic theory. *Am Math Mon* 99(1):31–41
- Friedman N, King J (1991) Rank one lightly mixing. *Israel J Math* 73:281–288
- Friedman NA, Ornstein DS (1970) On isomorphism of weak Bernoulli transformations. *Adv Math* 5:365–394
- Friedman NA, Ornstein DS (1973) Ergodic transformations induce mixing transformations. *Adv Math* 10(1):147–163
- Furstenberg H (1977) Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J Analyse Math* 31:204–256
- Furstenberg H (1981) *Recurrence in ergodic theory and combinatorial number theory*. Princeton
- Furstenberg H, Katznelson Y (1978) An ergodic Szemerédi theorem for commuting transformations. *J d'Analyse Math* 34(1):275–291
- Furstenberg H, Weiss B (1978) The finite multipliers of infinite ergodic transformations. In: *The structure of*

- attractors in dynamical systems (Proc. Conf., North Dakota State Univ., Fargo, N.D., 1977). Springer
- Gao S, Hill A (2014) A model for rank one measure preserving transformations. *Topol Appl* 174:25–40
- Garsia AM (1965) A simple proof of E. Hopf's maximal ergodic theory. *J Math Mech* 14:381–382
- Góra P (1994) Properties of invariant measures for piecewise expanding one-dimensional transformations with summable oscillations of derivative. *Ergodic Theory Dynam Syst* 14:475–492
- Green B, Tao T (2008) The primes contain arbitrarily long arithmetic progressions. *Ann Math* 167(2):481–547
- Grivaux S (2013) IP-Dirichlet measures and IP-rigid dynamical systems: an approach via generalized riesz products. *Stud Math* 215(3):237–259
- Grivaux S, Roginskaya M (2013) Some new examples of recurrence and non-recurrence sets for products of rotations on the unit circle. *Czechoslov Math J* 63(3):603–627
- Halmos PA (1944) In general a measure-preserving transformation is mixing. *Ann Math* 45:786–792
- Halmos P (1956) *Lectures on ergodic theory*. Chelsea
- Halmos P, von Neumann J (1942) Operator methods in classical mechanics II. *Ann. Math* 43:332–350
- Hoffman C, Hecklen D (2002) Rational maps are d-adic Bernoulli. *Ann Math* 156:103–114
- Hoffman C, Rudolph DJ (2002) Uniform endomorphisms which are isomorphic to a Bernoulli shift. *Ann Math* 76:79–101
- Host B (1991) Mixing of all orders and pairwise independent joinings of systems with singular spectrum. *Israel J Math* 76:289–298
- Hu H (2004) Decay of correlations for piecewise smooth maps with indifferent fixed points. *Ergodic Theory Dynam Syst* 24:495–524
- Impagliazzo R, Kabanets V (2010) Constructive proofs of concentration bounds. In: Serna M, Shaltiel R, Jansen K, Rolim J (eds) *Approximation, randomization, and combinatorial optimization. algorithms and techniques*. Springer, Berlin, Heidelberg, pp 617–631
- Jewett RI (1970) The prevalence of uniquely ergodic systems. *J Math Mechanics* 19(8):717–729
- Kakutani S (1943) 131. Induced measure preserving transformations. *Proc Imperial Acad* 19(10):635–641
- Kakutani S (1986) *Selected papers*. Birkhauser
- Kalikow S (1982) T, T^{-1} transformation is not loosely Bernoulli. *Ann Math* 115:393–409
- Kalikow S (1984) Twofold mixing implies threefold mixing for rank one transformations. *Ergodic Theory Dynam Syst* 2:237–259
- Kamae T (1982) A simple proof of the ergodic theorem using non-standard analysis. *Israel J Math* 42:284–290
- Kammeyer JW (1990) A complete classification of the two-point extensions of a multidimensional Bernoulli shift. *J d Analyse Math* 54(1):113–163
- Katok A (1980a) Interval exchange transformations and some special flows are not mixing. *Israel J Math* 35:301–310
- Katok A (1980b) Smooth non-Bernoulli K-automorphisms. *Invent Math* 61:291–299
- Katok A, Hasselblatt B (1995) *Introduction to the modern theory of dynamical systems*, Cambridge
- Katok A, Stepin AM (1967) Approximations in ergodic theory. *Uspekhi Mat Nauk* 22:5(137):81–106
- Katznelson Y (1971) Ergodic automorphisms of T^n are Bernoulli shifts. *Israel J Math* 10:186–195
- Katznelson Y, Weiss B (1982) A simple proof of some ergodic theorems. *Israel J Math* 42:291–296
- Keane MS, Petersen KE (2006) Nearly simultaneous proofs of the ergodic theorem and maximal ergodic theorem. In: *Dynamics and stochastics: Festschrift in Honor of M. S. Keane*. Institute of Mathematical Statistics, pp 248–251
- Keane M, Smorodinsky M (1979a) Bernoulli schemes of the same entropy are finitarily isomorphic. *Ann Math* 109:397–406
- Keane M, Smorodinsky M (1979b) Finitary isomorphisms of irreducible markov shifts. *Israel J Math* 34(4):281–286
- Kolmogorov AN (1958) New metric invariant of transitive dynamical systems and endomorphisms of Lebesgue spaces. *Dokl Russ Acad Sci* 119:861–864
- Koopman BO (1931) Hamiltonian systems and Hilbert space. *Proc Nat Acad Sci* 17:315–218
- Kramli A, Simányi N, Szász D (1991) The K-property of three billiard balls. *Ann Math* 133:37–72
- Krieger W et al (1972) On unique ergodicity. In: *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability*, volume 2: Probability theory. The Regents of the University of California
- Le AN (2017) Nilsequences and multiple correlations along subsequences. *Ergodic Theory Dynam Syst*:1–21
- Ledrappier F (1978) Un champ Markovien peut être d'entropie nulle et mélangeant. *C R Acad Sci Paris Sér A-B* 287:561–563
- Lehrer E (1987) Topological mixing and uniquely ergodic systems. *Israel J Math* 57(2):239–255
- Lemanczyk M, Kanigowski A (2022) Spectral theory of dynamical systems. *Encyclopedia of Complexity and Systems Science*
- Liverani C (2004) Decay of correlations. *Ann Math* 159:1275–1312
- Marcus B, Petersen K, Weissman T (2011) *Entropy of hidden Markov processes and connections to dynamical systems*. London Mathematical Society
- Maruyama G (1949) *The harmonic analysis of stationary stochastic processes*. Memoirs of the Faculty of Science, Kyushu University. Series A Math 4(1):45–106
- Masser DW (2004) Mixing and linear equations over groups in positive characteristic. *Israel J Math* 142:189–204
- Masur H (1982) Interval exchange transformations and measured foliations. *Ann Math* 115:169–200
- McCutcheon R, Quas A (2007) Generalized polynomials and mild mixing systems. *Can J Math*. to appear
- McGoff ANK, Mukherjee S, Pillai N (2015) Consistency of maximum likelihood estimation for some dynamical systems. *Ann Stat* 43(1):1–29
- Méla X, Petersen K (2005) Dynamical properties of the Pascal adic transformation. *Ergodic Theory Dynam Syst* 25:227–256

- Muehlegger E, Raich A, Silva C, Zhao W (1997) Lightly mixing on dense algebras. *Real Anal Exchange* 23: 259–265
- Newton D, Parry W (1966) On a factor automorphism of a normal dynamical system. *Ann Math Statist* 37: 1528–1533
- Ornstein DS (1970) Bernoulli shifts with the same entropy are isomorphic. *Adv Math* 4:337–352
- Ornstein DS (1972) On the root in ergodic theory. In: *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability* (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory. University California Press, pp 347–356
- Ornstein DS (1973a) An example of a Kolmogorov automorphism that is not a Bernoulli shift. *Adv Math* 10: 49–62
- Ornstein DS (1973b) A K-automorphism with no square root and Pinsker’s conjecture. *Adv Math* 10:89–102
- Ornstein DS (1973c) A mixing transformation for which Pinsker’s conjecture fails. *Adv Math* 10:103–123
- Ornstein DS (1974) *Ergodic theory, randomness, and dynamical systems*. Yale University Press
- Ornstein DS, Shields PC (1973) An uncountable family of K-automorphisms. *Adv Math* 10:89–102
- Ornstein DS, Weiss B (1973) Geodesic flows are Bernoullian. *Israel J Math* 14:184–198
- Ornstein DS, Weiss B (1975) Unilateral codings of Bernoulli systems. *Israel J Math* 21:159166
- Oxtoby JC (1952) Ergodic sets. *Bull Amer Math Soc* 58: 116–136
- Parry W (1981) *Topics in ergodic theory*, Cambridge
- Pelekis C, Ramon J (2017) Hoeffding’s inequality for sums of dependent random variables. *Mediterr J Math* 14(6):243
- Petersen K (1983) *Ergodic theory*. Cambridge
- Petersen K, Schmidt K (1997) Symmetric Gibbs measures. *Trans Am Math Soc* 349:2775–2811
- Phelps R (1966) *Lectures on Choquet’s theorem*. Van Nostrand
- Pinsker MS (1960) Dynamical systems with completely positive or zero entropy. *Soviet Math Dokl* 1:937–938
- Pollard D (1990) Empirical processes: theory and applications. In: *NSF-CBMS regional conference series in probability and statistics*. JSTOR, pp 1–86
- Quas A (1996a) A C^1 expanding map of the circle which is not weak-mixing. *Israel J Math* 93:359–372
- Quas A (1996b) Non-ergodicity for C^1 expanding maps and g-measures. *Ergodic Theory Dynam Syst* 16: 531–543
- Raginsky M, Rakhlin A, Telgarsky M (2017) Non-convex learning via stochastic gradient langevin dynamics: a nonasymptotic analysis. In: *Kale S, Shamir O (eds) Proceedings of the 2017 Conference on learning theory*, volume 65 of *Proceedings of machine learning research*. PMLR, Amsterdam, pp 1674–1703
- Rényi A (1957) Representations for real numbers and their ergodic properties. *Acta Math Acad Sci Hungar* 8: 477–493
- Riesz F (1938) Some mean ergodic theorems. *J London Math Soc* 13:274–278
- Robertson D (2019) Mild mixing of certain interval-exchange transformations. *Ergodic Theory Dynam Syst* 39(1):248–256
- Rokhlin VA (1948) A ‘general’ measure-preserving transformation is not mixing. *Dokl Akad Nauk SSSR Ser Mat* 60:349–351
- Rokhlin VA (1949) On endomorphisms of compact commutative groups. *Izvestiya Akad Nauk SSSR Ser Mat* 13:329–340
- Rokhlin VA, Sinai Y (1961) Construction and properties of invariant measurable partitions. *Dokl Akad Nauk SSSR* 141:1038–1041
- Rudin W (1966) *Real and complex analysis*. McGraw Hill
- Rudolph D (1976) A two-valued step coding for ergodic flows. *Math Z* 150(3):201–220
- Rudolph DJ (1978) If a two-point extension of a Bernoulli shift has an ergodic square, then it is Bernoulli. *Israel J Math* 30(1–2):159–180
- Rudolph DJ (1990) *Fundamentals of measurable dynamics*. Oxford
- Ryzhikov VV (1993) Joinings and multiple mixing of the actions of finite rank. *Funct Anal Appl* 27:128–140
- Schmidt JP, Siegel A, Srinivasan A (1993) Chernoff-hoeffding bounds for applications with limited independence. In: *Proceedings of the fourth annual ACM-SIAM symposium on discrete algorithms, SODA 93*. Society for Industrial and Applied Mathematics, p 331340
- Shields P, Thouvenot J-P (1975) Entropy zero $\times$ Bernoulli processes are closed in the $\bar{d}$ – metric. *Ann Probab* 3: 732–736
- Silva CE (2008) *Invitation to ergodic theory*. Am Math Soc 42
- Simányi N (2003) Proof of the Boltzmann-Sinai ergodic hypothesis for typical hard disk systems. *Invent Math* 154:123–178
- Simányi N (2004) Proof of the ergodic hypothesis for typical hard ball systems. *Ann Henri Poincare* 5:203–233
- Simányi N, Szász D (1999) Hard ball systems are completely hyperbolic. *Ann Math* 149:35–96
- Sinai YG (1959) On the notion of entropy of a dynamical system. *Dokl Russ Acad Sci* 124:768–771
- Sinai YG (1964) On a weak isomorphism of transformations with invariant measure. *Mat Sb (NS)* 63:23–42
- Sinai YG (1970) Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. *Uspehi Mat Nauk* 25:141–192
- Sinai Y (1976) *Introduction to ergodic theory*. Princeton. translation of the 1973 Russian original
- Sinai YG, Chernov NI (1987) Ergodic properties of some systems of two-dimensional disks and three-dimensional balls. *Uspekhi Mat Nauk* 42:153–174
- Smorodinsky M (1971) A partition on a Bernoulli shift which is not weakly Bernoulli. *Math Syst Th* 5: 201–203
- Tao T (2008) Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory Dynam Syst* 28(2):657–688
- Tikhonov SV (2007) A complete metric in the set of mixing transformations. *Mat Sb* 198(4):135–158

- Vapnik V (2013) *The nature of statistical learning theory*. Springer
- Vapnik VN, Chervonenkis AY (2015) On the uniform convergence of relative frequencies of events to their probabilities. Springer, Cham, pp 11–30
- Veech W (1982) Gauss measures for transformations on the space on interval exchange maps. *Ann Math* 115: 201–242
- Vershik A (1974) A description of invariant measures for actions of certain infinite-dimensional groups. *Soviet Math Dokl* 15:1396–1400
- Vershik A (1981) Uniform algebraic approximation of shift and multiplication operators. *Soviet Math Dokl* 24: 97–100
- Vershik A, Livshits A (1992) Adic models of ergodic transformations, spectral theory, substitutions, and related topics. *Rep Theory Dynam Syst* 9:185–204
- von Neumann J (1932) Proof of the quasi-ergodic hypothesis. *Proc Natl Acad Sci U S A* 18:70–82
- Walters P (1972) Some invariant sigma algebras for measure preserving transformations. *Trans Am Math Soc* 163:357–368
- Walters P (1982) *An introduction to ergodic theory*. Springer
- Williams D (1991) *Probability with martingales*. Cambridge
- Young L-S (1995) Ergodic theory of differentiable dynamical systems. In: *Real and complex dynamical systems* (Hillerød, 1993). Kluwer, pp 293–336
- Young L-S (1998) Statistical properties of dynamical systems with some hyperbolicity. *Ann Math* 147:585–650
- Yuzvinskii SA (1967) Metric automorphisms with a simple spectrum. *Soviet Math. Dokl.* 8:243–245
- Zhang Q (1996) On convergence of the averages. *Monatshefte für Mathematik* 122(3):275–300

Ergodic Theory: Recurrence

Nikos Frantzikinakis¹ and Randall McCutcheon²

¹Department of Mathematics, University of Crete, Heraklion, Greece

²Department of Mathematics, University of Memphis, Memphis, TN, USA

Article Outline

Glossary

Definition of the Subject and Its Importance

Introduction

Quantitative Poincaré Recurrence

Subsequence Recurrence

Multiple Recurrence

Connections with Combinatorics and Number

Theory

Future Directions

References

Glossary

Almost every, essentially Given a Lebesgue measure space $(X, \mathcal{B}, \mu)$, a property $P(x)$ predicated of elements of X is said to hold for almost every $x \in X$, if the set $X \setminus \{x: P(x) \text{ holds}\}$ has zero measure. Two sets $A, B \in \mathcal{B}$ are essentially disjoint if $\mu(A \cap B) = 0$.

Conservative system Is an infinite measure preserving system such that for no set $A \in \mathcal{B}$ with positive measure are $A, T^{-1}A, T^{-2}A, \dots$ pairwise essentially disjoint.

(c_n) -Conservative system If $(c_n)_{n \in \mathbb{N}}$ is a decreasing sequence of positive real numbers, a conservative ergodic measure preserving transformation T is (c_n) -conservative if for some nonnegative function $f \in L^1(\mu)$, $\sum_{n=1}^{\infty} c_n f(T^n x) = \infty$ a.e.

Doubling map If $\mathbb{T}$ is the interval $[0, 1]$ with its endpoints identified and addition performed modulo 1, the (non-invertible) transformation

$T: \mathbb{T} \rightarrow \mathbb{T}$, defined by $Tx = 2x \bmod 1$, preserves Lebesgue measure, hence induces a measure preserving system on $\mathbb{T}$.

Ergodic system Is a measure preserving system $(X, \mathcal{B}, \mu, T)$ (finite or infinite) such that every $A \in \mathcal{B}$ that is T -invariant (i.e., $T^{-1}A = A$) satisfies either $\mu(A) = 0$ or $\mu(X \setminus A) = 0$. (One can check that the rotation R_α is ergodic if and only if α is irrational and that the doubling map is ergodic.)

Ergodic decomposition Every measure-preserving system $(X, \mathcal{X}, \mu, T)$ can be expressed as an integral of ergodic systems; for example, one can write $\mu = \int \mu_t d\lambda(t)$, where λ is a probability measure on $[0, 1]$ and μ_t are T -invariant probability measures on $(X, \mathcal{X})$ such that the systems $(X, \mathcal{X}, \mu_t, T)$ are ergodic for $t \in [0, 1]$.

Ergodic theorem States that if $(X, \mathcal{B}, \mu, T)$ is a measure preserving system and $f \in L^2(\mu)$, then $\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N T^n f - P_f \right\|_{L^2(\mu)} = 0$, where P_f denotes the orthogonal projection of the function f onto the subspace $\{f \in L^2(\mu): T f = f\}$.

Hausdorff a -measure Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system endowed with a μ -compatible metric d . The Hausdorff a -measure $\mathcal{H}_a(X)$ of X is an outer measure defined for all subsets of X as follows: First, for $A \subset X$ and $\varepsilon > 0$, let $\mathcal{H}_{a,\varepsilon}(A) = \inf \left\{ \sum_{i=1}^{\infty} r_i^a \right\}$, where the infimum is taken over all countable coverings of A by sets $U_i \subset X$ with diameter $r_i < \varepsilon$. Then define $H_a(A) = \limsup_{\varepsilon \rightarrow 0} \mathcal{H}_{a,\varepsilon}(A)$.

Infinite measure-preserving system Same as measure preserving system, but $\mu(X) = \infty$.

Invertible system Is a measure-preserving system $(X, \mathcal{B}, \mu, T)$ (finite or infinite), with the property that there exists $X_0 \in \mathcal{B}$, with $\mu(X \setminus X_0) = 0$, and such that the transformation $T: X_0 \rightarrow X_0$ is bijective, with T^{-1} measurable.

Measure-preserving system Is a quadruple $(X, \mathcal{B}, \mu, T)$, where X is a set, $\mathcal{B}$ is a σ -algebra of subsets of X (i.e., $\mathcal{B}$ is closed under countable

unions and complementation), μ is a probability measure (i.e., a countably additive function from $\mathcal{B}$ to $[0, 1]$ with $\mu(X) = 1$), and $T: X \rightarrow X$ is measurable (i.e., $T^{-1}A = \{x \in X: T x \in A\} \in \mathcal{B}$ for $A \in \mathcal{B}$) and μ -preserving (i.e., $\mu(T^{-1}A) = \mu(A)$). Moreover, throughout the discussion, we assume that the measure space $(X, \mathcal{B}, \mu)$ is Lebesgue (see Section 1.0 of (Aronson 1997)).

μ -Compatible metric Is a separable metric on X , where $(X, \mathcal{B}, \mu)$ is a probability space, having the property that open sets measurable.

Positive definite sequence Is a complex-valued sequence $(a_n)_{n \in \mathbb{Z}}$ such that for any $z_1, \dots, z_k \in \mathbb{C}$, $\sum_{i,j=1}^k a_{i-j} z_i \bar{z}_j \geq 0$.

Rotations on $\mathbb{T}$ If $\mathbb{T}$ is the interval $[0, 1]$ with its endpoints identified and addition performed modulo 1, then for every $\alpha \in \mathbb{R}$, the transformation $R_\alpha: \mathbb{T} \rightarrow \mathbb{T}$, defined by $R_\alpha x = x + \alpha$, preserves Lebesgue measure on $\mathbb{T}$ and hence induces a measure preserving system on $\mathbb{T}$.

Syndetic set Is a subset $E \subset \mathbb{Z}$ having bounded gaps. If G is a general discrete group, a set $E \subset G$ is syndetic if $G = F E$ for some finite set $F \subset G$.

Upper density Is the number $\bar{d}(\Lambda) = \limsup_{N \rightarrow \infty} \frac{|\Lambda \cap \{-N, \dots, N\}|}{2N+1}$, where $\Lambda \subset \mathbb{Z}$ (assuming the limit to exist). Alternatively for measurable $E \subset \mathbb{R}^m$, $\bar{d}(E) = \limsup_{l(S) \rightarrow \infty} \frac{m(S \cap E)}{m(S)}$, where S ranges over all cubes in $\mathbb{R}^m$ and $l(S)$ denotes the length of the shortest edge of S .

Notation The following notation will be used throughout the article: $Tf = f \circ T$, $\{x\} = x - [x]$, $D\text{-}\lim_{n \rightarrow \infty} (a_n) = a \Leftrightarrow \bar{d}(\{n: |a_n - a| > \varepsilon\}) = 0$ for every $\varepsilon > 0$.

Definition of the Subject and Its Importance

The basic principle that lies behind several recurrence phenomena is that the typical trajectory of a system with finite volume comes back infinitely often to any neighborhood of its initial point. This

principle was first exploited by Poincaré in his 1890 King Oscar prize-winning memoir that studied planetary motion. Using the prototype of an ergodic theoretic argument, he showed that in any system of point masses having fixed total energy that restricts its dynamics to bounded subsets of its phase space, the typical state of motion (characterized by configurations and velocities) must recur to an arbitrary degree of approximation.

Among the recurrence principle's more spectacularly counterintuitive ramifications is that isolated ideal gas systems that do not lose energy will return arbitrarily closely to their initial states, even when such a return entails a decrease in entropy from equilibrium, in apparent contradiction to the second law of thermodynamics. Such concerns, previously canvassed by Poincaré himself, were more infamously expounded by Zermelo (1896) in 1896. Subsequent clarifications by Boltzmann, Maxwell, and others led to an improved understanding of the second law's primarily statistical nature. (For an interesting historical/philosophical discussion, see Sklar (2004) and also Bergelson (2000). For a probabilistic analysis of the likelihood of observing second law violations in small systems over short time intervals, see Evans and Searles (2002).)

These discoveries had a profound impact in dynamics, and the theory of measure-preserving transformations (ergodic theory) evolved from these developments. Since then, the Poincaré recurrence principle has been applied to a variety of different fields in mathematics, physics, and information theory. In this article we survey the impact it has had in ergodic theory, especially as pertains to the field of *ergodic Ramsey theory*. (The heavy emphasis herein on the latter reflects authorial interest and is not intended to transmit a proportionate image of the broader landscape of research relating to recurrence in ergodic theory.) Background information we assume in this article can be found in the books (Einsiedler and Ward 2011; Furstenberg 1981; Glasner 2003; Host and Kra 2018; Petersen 1989; Walters 1982) (see also the chapter "Measure Preserving Systems" by

K. Petersen in this volume). Related information can also be found on the survey articles (Bergelson 1996, 2006a, b; Frantzikinakis 2016; Kra 2006b, 2007, 2011).

Introduction

In this section we shall give several formulations of the Poincaré recurrence principle using the language of ergodic theory. Roughly speaking, the principle states that in a finite (or conservative) measure-preserving system, every set of positive measure (or almost every point) comes back to itself infinitely many times under iteration. Despite the profound importance of these results, their proofs are extremely simple.

Theorem 2.1 (Poincaré Recurrence for Sets)

Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and $A \in \mathcal{B}$ with $\mu(A) > 0$. Then $\mu(A \cap T^{-n}A) > 0$ for infinitely many $n \in \mathbb{N}$.

Proof Since T is measure preserving, the sets $A, T^{-1}A, T^{-2}A, \dots$ have the same measure. These sets cannot be pairwise essentially disjoint, since then the union of finitely many of them would have measure greater than $\mu(X) = 1$. Therefore, there exist $m, n \in \mathbb{N}$, with $n > m$, such that $\mu(T^{-m}A \cap T^{-n}A) > 0$. Again since T is measure preserving, we conclude that $\mu(A \cap T^{-k}A) > 0$, where $k = n - m > 0$. Repeating this argument for the iterates $A, T^{-m}A, T^{-2m}A, \dots$, for all $m \in \mathbb{N}$, we easily deduce that $\mu(A \cap T^{-n}A) > 0$ for infinitely many $n \in \mathbb{N}$.

We remark that the above argument actually shows that $\mu(A \cap T^{-n}A) > 0$ for some $n \leq \left\lceil \frac{1}{\mu(A)} \right\rceil + 1$.

Theorem 2.2 (Poincaré Recurrence for Points)

Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and $A \in \mathcal{B}$. Then for almost every $x \in A$, we have that $T^n x \in A$ for infinitely many $n \in \mathbb{N}$.

Proof Let B be the set of $x \in A$ such that $T^n x \notin A$ for all $n \in \mathbb{N}$. Notice that $B = A \setminus \bigcap_{n \in \mathbb{N}} T^{-n}A$; in particular, B is measurable. Since the iterates $B, T^{-1}B, T^{-2}B, \dots$ are pairwise essentially disjoint,

we conclude (as in the proof of Theorem 2.1) that $\mu(B) = 0$. This shows that for almost every $x \in A$, we have that $T^n x \in A$ for some $n \in \mathbb{N}$. Repeating this argument for the transformation T^m in place of T for all $m \in \mathbb{N}$, we easily deduce the advertised statement.

Next we give a variation of Poincaré recurrence for measure-preserving systems endowed with a compatible metric.

Theorem 2.3 (Poincaré Recurrence for Metric Systems)

Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system, and suppose that X is endowed with a μ -compatible metric. Then for almost every $x \in X$, we have

$$\liminf_{n \rightarrow \infty} d(x, T^n x) = 0.$$

The proof of this result is similar to the proof of Theorem 2.2 (see Furstenberg 1981, p. 61). Applying this result to the doubling map $T x = 2x$ on $\mathbb{T}$, we get that for almost every $x \in X$, every string of zeros and ones in the dyadic expansion of x occurs infinitely often.

We remark that all three formulations of the Poincaré Recurrence Theorem that we have given hold for conservative systems as well. See, e.g., Aaronson (1997) for details.

This article is structured as follows. In section “Quantitative Poincaré Recurrence,” we give a few quantitative versions of the previously mentioned qualitative results. In sections “Subsequence Recurrence” and “Multiple Recurrence,” we give several refinements of the Poincaré recurrence theorem, by restricting the scope of the return time n and by considering multiple intersections (for simplicity we focus on $\mathbb{Z}$ -actions). In section “Connections with Combinatorics and Number Theory,” we give various implications of the recurrence results in combinatorics and number theory (see also the chapter ► “Ergodic Theory: Interactions with Combinatorics and Number Theory” by T. Ward in the present volume). Lastly, in section “Future Directions,” we give several open problems related to the material

presented in sections “[Subsequence Recurrence](#),” “[Multiple Recurrence](#),” and “[Connections with Combinatorics and Number Theory](#).”

Quantitative Poincaré Recurrence

Early Results

For applications it is desirable to have quantitative versions of the results mentioned in the previous section. For example one would like to know how large $\mu(A \cap T^{-n}A)$ can be made and for how many n .

Theorem 3.1 (Khinchine 1934) *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and $A \in \mathcal{B}$. Then for every $\varepsilon > 0$, we have $\mu(A \cap T^{-n}A) > \mu(A)^2 - \varepsilon$ for a set of $n \in \mathbb{N}$ that has bounded gaps.*

By considering the doubling map $Tx = 2x$ on $\mathbb{T}$ and letting $A = \mathbf{1}_{[0, 1/2]}$, it is easy to check that the lower bound of the previous result cannot be improved. We also remark that it is not possible to estimate the size of the gap by a function of $\mu(A)$ alone. One can see this by considering the rotations $R_k x = x + 1/k$ for $k \in \mathbb{N}$, defined on $\mathbb{T}$, and letting $A = \mathbf{1}_{[0, 1/3]}$.

Concerning the second version of the Poincaré recurrence theorem, it is natural to ask whether for almost every $x \in X$ the set of return times $S_x = \{n \in \mathbb{N} : T^n x \in A\}$ has bounded gaps. This is not the case, as one can see by considering the doubling map $Tx = 2x$ on $\mathbb{T}$ with the Lebesgue measure and letting $A = \mathbf{1}_{[0, 1/2]}$. Since Lebesgue almost every $x \in \mathbb{T}$ contains arbitrarily large blocks of ones in its dyadic expansion, the set S_x has unbounded gaps. Nevertheless, as an easy consequence of the Birkhoff ergodic theorem (Birkhoff 1931), one has the following.

Theorem 3.2 *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and $A \in \mathcal{B}$ with $\mu(A) > 0$. Then for almost every $x \in X$, the set $S_x = \{n \in \mathbb{N} : T^n x \in A\}$ has well-defined density and $\int d(S_x)$*

$d\mu(x) = \mu(A)$. Furthermore, for ergodic measure-preserving systems, we have $d(S_x) = \mu(A)$ a.e.

Another question that arises naturally is, given a set A with positive measure and an $x \in A$, how long should one wait till some iterate $T^n x$ of x hits A . By considering an irrational rotation R_α on $\mathbb{T}$, where α is very near to, but not less than, $\frac{1}{100}$, and letting $A = \mathbf{1}_{[0, 1/2]}$, one can see that the first return time is a member of the set $\{1, 50, 51\}$. So it may come as a surprise that the average first return time does not depend on the system (as long as it is ergodic) but only on the measure of the set A .

Theorem 3.3 (Kac 1947) *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure-preserving system and $A \in \mathcal{B}$ with $\mu(A) > 0$. For $x \in X$ define $R_A(x) = \min\{n \in \mathbb{N} : T^n x \in A\}$. Then for $x \in A$ the expected value of $R_A(x)$ is $1/\mu(A)$, i.e., $\int_A R_A(x) d\mu = 1$.*

More Recent Results

As we mentioned in the previous section, if the space X is endowed with a μ -compatible metric d , then for almost every $x \in X$, we have that $\liminf_{n \rightarrow \infty} d(x, T^n x) = 0$. A natural question is how much iteration is needed to come back within a small distance of a given typical point. Under some additional hypothesis on the metric d , we have the following answer.

Theorem 3.4 (Boshernitzan 1993) *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system endowed with a μ -compatible metric d . Assume that the Hausdorff μ -measure $\mathcal{H}_\mu(X)$ of X is σ -finite (i.e., X is a countable union of sets X_i with $\mathcal{H}_\mu(X_i) < \infty$). Then for almost every $x \in X$*

$$\liminf_{n \rightarrow \infty} \left\{ n^{\frac{1}{\mu}} \cdot d(x, T^n x) \right\} < \infty.$$

Furthermore, if $\mathcal{H}_\mu(X) = 0$, then for almost every $x \in X$

$$\liminf_{n \rightarrow \infty} \left\{ n^{\frac{1}{\mu}} \cdot d(x, T^n x) \right\} = 0.$$

One can see from rotations by “badly approximable” vectors $\alpha \in \mathbb{T}^k$ that the exponent $1/k$ in the previous theorem cannot be improved. Several applications of Theorem 3.4 to billiard flows, dyadic transformations, symbolic flows, and interval exchange transformations are given in Boshernitzan (1993). For a related result dealing with mean values of the limits in Theorem 3.4, see Shkredov (2002).

An interesting connection between rates of recurrence and entropy of an ergodic measure-preserving system was established by Ornstein and Weiss (1993), following earlier work of Wyner and Ziv (1989).

Theorem 3.5 (Ornstein and Weiss 1993) *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure-preserving system and $\mathcal{P}$ be a finite partition of X . Let $P_n(x)$ be the element of the partition $\bigvee_{i=0}^{n-1} T^{-i} \mathcal{P} = \{\bigcap_{i=0}^{n-1} T^{-i} P^{(i)} : P^{(i)} \in \mathcal{P}, 0 \leq i < n\}$ that contains x . Then for almost every $x \in X$, the first return time $R_n(x)$ of x to $P_n(x)$ is asymptotically equivalent to $e^{h(T, \mathcal{P})n}$, where $h(T, \mathcal{P})$ denotes the entropy of the system with respect to the partition $\mathcal{P}$. More precisely,*

$$\lim_{n \rightarrow \infty} \frac{\log R_n(x)}{n} = h(T, \mathcal{P}).$$

An extension of the above result to some classes of infinite measure-preserving systems was given in Galatolo et al. (2006).

Another connection of recurrence rates, this time with the local dimension of an invariant measure, is given by the next result.

Theorem 3.6 (Barreira 2001) *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure-preserving system. Define the upper and lower recurrence rates*

$$\underline{R}(x) = \liminf_{r \rightarrow 0} \frac{\log \tau_r(x)}{-\log r} \text{ and } \bar{R}(x) = \limsup_{r \rightarrow 0} \frac{\log \tau_r(x)}{-\log r},$$

where $\tau_r(x)$ is the first return time of $T^k x$ in $B(x, r)$ and the upper and lower pointwise dimensions

$$\underline{d}_\mu(x) = \liminf_{r \rightarrow 0} \frac{\log \mu(B(x, r))}{\log r} \text{ and } \bar{d}_\mu(x) = \limsup_{r \rightarrow 0} \frac{\log \mu(B(x, r))}{\log r}.$$

Then for almost every $x \in X$, we have

$$\underline{R}(x) \leq \underline{d}_\mu(x) \text{ and } \bar{R}(x) \leq \bar{d}_\mu(x).$$

Roughly speaking, this theorem asserts that for typical $x \in X$ and for small r , the first return time of x in $B(x, r)$ is at most $r^{-d_\mu(x)}$. Since $\underline{d}_\mu(x) \leq \mathcal{H}_a(X)$ for almost every $x \in X$, we can conclude the first part of Theorem 3.4 from Theorem 3.6. For related results the interested reader should consult the survey (Barreira 2005) and the bibliography therein.

We also remark that the previous results and related concepts have been applied to estimate the dimension of certain strange attractors (see Hasley and Jensen (2004) and the references therein) and the entropy of certain Gibbsian systems (Chazottes and Ugalde 2005).

We end this section with a result that connects “wandering rates” of sets in infinite measure-preserving systems with their “recurrence rates.” The next theorem follows easily from a result about lower bounds on ergodic averages for measure-preserving systems due to Leibman (2002); a weaker form for conservative, ergodic systems can be found in Aaronson (1981).

Theorem 3.7 *Let $(X, \mathcal{B}, \mu, T)$ be an infinite measure-preserving system and $A \in \mathcal{B}$ with $\mu(A) < \infty$. Then for all $N \in \mathbb{N}$,*

$$\left(\frac{\mu\left(\bigcup_{n=0}^{N-1} T^{-n} A\right)}{N} \cdot \sum_{n=0}^{N-1} \mu(A \cap T^{-n} A) \right) \geq \frac{1}{2} \cdot (\mu(A))^2.$$

Subsequence Recurrence

In this section we discuss what restrictions we can impose on the set of return times in the various versions of the Poincaré recurrence theorem. We start with:

Definition 4.1 Let $R \subset \mathbb{Z}$. Then R is a set of:

- (a) *Recurrence* if for any invertible measure-preserving system $(X, \mathcal{B}, \mu, T)$ and $A \in \mathcal{B}$ with $\mu(A) > 0$, there is some nonzero $n \in R$ such that $\mu(A \cap T^{-n}A) > 0$.
- (b) *Topological recurrence* if for every compact metric space (X, d) , continuous transformation $T: X \rightarrow X$ and every $\varepsilon > 0$, there are $x \in X$ and nonzero $n \in R$ such that $d(x, T^n x) < \varepsilon$.

It is easy to check that the existence of a single $n \in R$ satisfying the previous recurrence conditions actually guarantees the existence of infinitely many $n \in R$ satisfying the same conditions. Moreover, if R is a set of recurrence then one can see from existence of some T -invariant measure μ that R is also a set of topological recurrence. A (complicated) example showing that the converse is not true was given by Kriz (1987) (see also Forrest (1991) and McCutcheon (1995) for simplified versions).

Before giving a list of examples of sets of (topological) recurrence, we discuss some necessary conditions: A set of topological recurrence must contain infinitely many multiples of every positive integer, as one can see by considering rotations on $\mathbb{Z}_d$, $d \in \mathbb{N}$. Hence, the sets $\{2n+1, n \in \mathbb{N}\}$, $\{n^2+1, n \in \mathbb{N}\}$, and $\{p+2, p \text{ prime}\}$ are not good for (topological) recurrence. If $(s_n)_{n \in \mathbb{N}}$ is a lacunary sequence (meaning $\liminf_{n \rightarrow \infty} (s_{n+1}/s_n) = \rho > 1$), then one can construct an irrational number α such that $\{s_n \alpha\} \in [\delta, 1 - \delta]$ for all large $n \in \mathbb{N}$, where $\delta > 0$ depends on ρ (see Katznelson (2001) for example). As a consequence, the sequence $(s_n)_{n \in \mathbb{N}}$ is not good for (topological) recurrence. Lastly, we mention that by

considering product systems, one can immediately show that any set of (topological) recurrence R is partition regular, meaning that if R is partitioned into finitely many pieces then at least one of these pieces must still be a set of (topological) recurrence. Using this observation, one concludes, for example, that any union of finitely many lacunary sequences is not a set of recurrence. We present now some examples of sets of recurrence.

Theorem 4.2 *The following are sets of recurrence:*

- (i) Any set of the form $\bigcup_{n \in \mathbb{N}} \{a_n, 2a_n, \dots, na_n\}$ where $a_n \in \mathbb{N}$. This follows from a finitary version of Szemerédi's theorem.
- (ii) Any IP-set, meaning a set that consists of all finite sums of some infinite set (Furstenberg and Katznelson 1985).
- (iii) Any difference set $S - S$, meaning a set that consists of all possible differences of some infinite set S .
- (iv) The set $\{p(n), n \in \mathbb{N}\}$ where p is any nonconstant integer polynomial with $p(0) = 0$ (Furstenberg 1981; Sárközy 1978). In fact we only have to assume that the range of the polynomial contains multiples of an arbitrary positive integer; this follows from Theorem 4.3 below.
- (v) The set $\{p(n), n \in S\}$, where p is an integer polynomial with $p(0) = 0$ and S is any IP-set (Bergelson et al. 1996).
- (vi) The set of values of an admissible generalized polynomial. This class contains in particular the smallest function algebra G containing all integer polynomials having zero constant term and such that if $g_1, \dots, g_k \in G$ and $c_1, \dots, c_k \in \mathbb{R}$, then $\sum_{i=1}^k c_i g_i \in G$, where $x = [x + \frac{1}{2}]$ denotes the integer nearest to x (Bergelson et al. 2006).
- (vii) The set of shifted primes $\{p-1, p \text{ prime}\}$, the set $\{p+1, p \text{ prime}\}$ (Sárközy 1978), and also the integer part of certain smooth functions evaluated at the primes (Bergelson et al. 2019).

- (viii) $R = \{[a(1)], [a(2)], \dots\}$, where $a(x) = x^c$ for any $c > 0$. This follows from Theorem 4.3 below and standard exponential sum estimates; see also Boshernitzan et al. (2005) for a more general result regarding Hardy sequences.
- (ix) The set of values of a random non-lacunary sequence. More precisely, pick $n \in \mathbb{N}$ independently with probability b_n where $0 \leq b_n \leq 1$ is decreasing and $\lim_{n \rightarrow \infty} nb_n = \infty$, then the resulting set is almost surely a set of recurrence (this follows from Bourgain (1988)).
- (x) Arbitrary shifts of integers with an even (or odd) number of prime factors and other similar sets with arithmetic structure (Bergelson et al. “A structure theorem for...”; Frantzikinakis and Host 2017b).

Showing that the first three sets are good for recurrence is a straightforward modification of the argument used to prove Theorem 2.1. The other examples require more work.

A criterion of Kamae and Mendés-France (1978) provides a powerful tool that may be used in many instances to establish that a set R is a set of recurrence. We mention a variation of their result.

Theorem 4.3 (Kamae and Mendés-France 1978) Suppose that $R = \{a_1 < a_2 < \dots\}$ is a subset of $\mathbb{N}$ such that:

- (i) The sequence $\{a_n \alpha\}_{n \in \mathbb{N}}$ is uniformly distributed in $\mathbb{T}$ for every irrational α .
- (ii) The set $R_m = \{n \in \mathbb{N} : m|a_n\}$ has positive upper density for every $m \in \mathbb{N}$.

Then R is a set of recurrence.

We sketch a proof for this result. First, recall Herglotz’s theorem: if $(a_n)_{n \in \mathbb{Z}}$ is a positive definite sequence, then there is a unique measure σ on the torus $\mathbb{T}$ such that $a_n = \int e^{2\pi i n t} d\sigma(t)$. The case of interest to us is $a_n = \int_{\mathbb{T}} f(x) f(T^n x) d\mu$, where

T is measure preserving and $f \in L^\infty(\mu)$; (a_n) is positive definite, and we call $\sigma = \sigma_f$ the spectral measure of f .

Let now $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and $A \in \mathcal{B}$ with $\mu(A) > 0$. Putting $f = \mathbf{1}_A$, one has

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int f(x) \cdot f(T^{a_n} x) d\mu \\ = \int_{\mathbb{T}} \lim_{N \rightarrow \infty} \left(\frac{1}{N} \sum_{n=1}^N e^{2\pi i a_n t} \right) d\sigma_f(t) \end{aligned} \quad (1)$$

where we assume for simplicity that all limits exist. For t irrational the limit inside the integral is zero (by condition (i)), so the last integral can be taken over the rational points in $\mathbb{T}$. Since the spectral measure of a function orthogonal to the subspace

$$\mathcal{H} = \overline{\{f \in L^2(\mu) : \text{there exists } k \in \mathbb{N} \text{ with } T^k f = f\}} \quad (2)$$

has no rational point masses, we can easily deduce that when computing the first limit in (1), we can replace the function f by its orthogonal projection g onto the subspace $\mathcal{H}$ (g is again nonnegative and $g \neq 0$). To complete the argument, we approximate g by a function g' such that $T^m g' = g'$ for some appropriately chosen m and use condition (ii) to deduce that the limit of the average (1) is positive.

In order to apply Theorem 4.3, one uses the standard machinery of uniform distribution. Recall Weyl’s criterion: a real valued sequence $(x_n)_{n \in \mathbb{N}}$ is uniformly distributed mod 1 if for every nonzero $k \in \mathbb{Z}$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N e^{2\pi i k x_n} = 0.$$

This criterion becomes especially useful when paired with van der Corput’s so-called third principal property: if, for every $h \in \mathbb{N}$, $(x_{n+h} - x_n)_{n \in \mathbb{N}}$ is uniformly distributed mod 1, then $(x_n)_{n \in \mathbb{N}}$ is

uniformly distributed mod 1. Using the foregoing criteria and some standard (albeit nontrivial) exponential sum estimates, one can verify, for example, that the sets (iv) and (vii) in Theorem 4.2 are good for recurrence.

In light of the connection elucidated above between uniform distribution mod 1 and recurrence, it is not surprising that van der Corput's method has been adapted by modern ergodic theorists for use in establishing recurrence properties directly.

Theorem 4.4 (Bergelson 1987a) *Let $(x_n)_{n \in \mathbb{N}}$ be a bounded sequence in a Hilbert space. If*

$$D - \lim_{m \rightarrow \infty} \left(\lim_{N \rightarrow \infty} \frac{1}{N} \langle x_{n+m}, x_n \rangle \right) = 0,$$

then

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N x_n \right\| = 0.$$

Let us illustrate how one uses this “van der Corput trick” by showing that $S = \{n^2 : n \in \mathbb{N}\}$ is a set of recurrence. We will actually establish the following stronger fact: If $(X, \mathcal{B}, \mu, T)$ is a measure-preserving system and $f \in L^\infty(\mu)$ is nonnegative and $f \neq 0$, then

$$\liminf_{n \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int f(x) \cdot f(T^{n^2} x) d\mu > 0. \quad (3)$$

Then our result follows by setting $f = \mathbf{1}_A$ for some $A \in \mathcal{B}$ with $\mu(A) > 0$. The main idea is one that occurs frequently in ergodic theory; split the function f into two components, one of which contributes zero to the limit appearing in (3) and the other one being much easier to handle than f . To do this consider the T -invariant subspace of $L^2(X)$ defined by

$$\mathcal{H} = \overline{\{f \in L^2(\mu) : \text{there exists } k \in \mathbb{N} \text{ with } T^k f = f\}}. \quad (4)$$

Write $f = g + h$ where $g \in \mathcal{H}$ and $h \perp \mathcal{H}$, and expand the average in (3) into a sum of four averages involving the functions g and h . Two of these averages vanish because iterates of g are orthogonal to iterates of h . So in order to show that the only contribution comes from the average that involves the function g alone, it suffices to establish that

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N T^{n^2} h \right\|_{L^2(\mu)} = 0. \quad (5)$$

To show this we will apply the Hilbert space van der Corput lemma. For given $h \in \mathbb{N}$, we let $x_n = T^n h$ and compute

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \langle x_{n+m}, x_n \rangle &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int T^{n^2+2nm+m^2} h \cdot T^{n^2} h d\mu \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int T^{2nm} (T^{m^2} h) \cdot h d\mu. \end{aligned}$$

Applying the von Neumann ergodic theorem (von Neumann 1932) to the transformation T^{2m} and using the fact that $h \perp \mathcal{H}$, we get that the last limit is 0. This implies (5).

Thus far we have shown that in order to compute the limit in (3), we can assume that $f = g \in \mathcal{H}$ (g is also nonnegative and $g \neq 0$). By the definition of $\mathcal{H}$, given any $\varepsilon > 0$, there exists a function $f' \in \mathcal{H}$ such that $T^k f' = f'$ for some $k \in \mathbb{N}$ and $\|f - f'\|_{L^2(\mu)} \leq \varepsilon$. Then the limit in (3) is at least $1/k$ times the limit

$$\liminf_{n \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int f(x) \cdot f(T^{(kn)^2} x) d\mu.$$

Applying the triangle inequality twice, we get that this is greater or equal than

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int f'(x) \cdot f'(T^{(kn)^2} x) d\mu - c \cdot \varepsilon &= \int (f'(x))^2 d\mu - 2\varepsilon \\ &\geq \left(\int f'(x) d\mu \right)^2 - c \cdot \varepsilon, \end{aligned}$$

for some constant c that does not depend on ε (we used that $T^k f' = f'$ and the Cauchy-Schwartz inequality). Choosing ε small enough, we

conclude that the last quantity is positive, completing the proof.

Multiple Recurrence

Simultaneous multiple returns of positive measure sets to themselves were first considered by H. Furstenberg (1977), who gave a new proof of Szemerédi's theorem (Szemerédi 1975) on arithmetic progressions by deriving it from the following theorem.

Theorem 5.1 (Furstenberg 1977) *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and $A \in \mathcal{B}$ with $\mu(A) > 0$. Then for every $k \in \mathbb{N}$, there is some $n \in \mathbb{N}$ such that*

$$\mu(A \cap T^{-n}A \cap \dots \cap T^{-kn}A) > 0. \quad (6)$$

Furstenberg's proof came by means of a new structure theorem allowing one to decompose an arbitrary measure-preserving system into component elements exhibiting one of two extreme types of behavior: *compactness*, characterized by regular, "almost periodic" trajectories, and *weak mixing*, characterized by irregular, "quasi-random" trajectories. On $\mathbb{T}$, these types of behavior are exemplified by rotations and by the doubling map, respectively. To see the point, imagine trying to predict the initial digit of the dyadic expansion of $T^n x$ given knowledge of the initial digits of $T^i x$, $1 \leq i < n$. We use the case $k = 2$ to illustrate the basic idea.

It suffices to show that if $f \in L^\infty(\mu)$ is nonnegative and $f \neq 0$, one has

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int f(x) \cdot f(T^n x) \cdot f(T^{2n} x) d\mu > 0. \quad (7)$$

An ergodic decomposition argument enables us to assume that our system is ergodic. As in the earlier case of the squares, we split f into "almost periodic" and "quasi-random"

components. Let K be the closure in L^2 of the subspace spanned by the eigenfunctions of T , i.e., the functions $f \in L^2(\mu)$ that satisfy $f(Tx) = e^{2\pi i \alpha} f(x)$ for some $\alpha \in \mathbb{R}$. We write $f = g + h$, where $g \in \mathcal{K}$ and $h \perp \mathcal{K}$. It can be shown that $g, h \in L^\infty(\mu)$ and g is again nonnegative with $g \neq 0$. We expand the average in (7) into a sum of eight averages involving the functions g and h . In order to show that the only nonzero contribution to the limit comes from the term involving g alone, it suffices to establish that

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N T^n g \cdot T^{2n} h \right\|_{L^2(\mu)} = 0, \quad (8)$$

(and similarly with h and g interchanged, and with $g = h$, which is similar). To establish (8), we use the Hilbert space van der Corput lemma on $x_n = T^n g \cdot T^{2n} h$. Some routine computations and a use of the ergodic theorem reduce the task to showing that

$$D - \lim_{m \rightarrow \infty} \left(\int h(x) \cdot h(T^{2m} x) d\mu \right) = 0.$$

But this is well known for $h \perp \mathcal{K}$ (e.g., in virtue of the fact that for $h \perp \mathcal{K}$ the spectral measure σ_h is continuous). We are left with the average (7) when $f = g \in \mathcal{K}$. In this case f can be approximated arbitrarily well by a linear combination of eigenfunctions, which easily implies that given $\varepsilon > 0$, one has $\|T^n f - f\|_{L^2(\mu)} \leq \varepsilon$ for a set of $n \in \mathbb{N}$ with bounded gaps. Using this fact and the triangle inequality, one finds that for a set of $n \in \mathbb{N}$ with bounded gaps,

$$\int f(x) \cdot f(T^n x) \cdot f(T^{2n} x) d\mu \geq \left(\int f d\mu \right)^3 - c \cdot \varepsilon$$

for a constant c that is independent of ε . Choosing ε small enough, we get (7).

The new techniques developed for the proof of Theorem 5.1 have led to a number of extensions, many of which have to date only ergodic proofs.

To expedite discussion of some of these developments, we introduce a definition:

Definition 5.2 Let $R \subset \mathbb{Z}$ and $k \in \mathbb{N}$. Then R is a set of k -recurrence if for every invertible measure-preserving system (X, B, μ, T) and $A \in B$ with $\mu(A) > 0$ there is some nonzero $n \in R$ such that

$$\mu(A \cap T^{-n}A \cap \dots \cap T^{-kn}A) > 0.$$

The notions of k -recurrence are distinct for different values of k . An example of a difference set that is a set of 1-recurrence but not a set of 2-recurrence was given in (Furstenberg 1977); sets of k -recurrence that are not sets of $(k+1)$ -recurrence for general k were given in Frantzikinakis et al. (2006) ($R_k = \{n \in \mathbb{N} : \{n^{k+1}\sqrt{2}\} \in [1/4, 3/4]\}$ is such).

Aside from difference sets, the sets of (1-)recurrence given in Theorem 4.2 may well be sets of k -recurrence for every $k \in \mathbb{N}$, though this has not been verified in all cases. Let us summarize the current state of knowledge. The following are sets of k -recurrence for every k : sets of the form $\cup_{n \in \mathbb{N}} \{a_n, 2a_n, \dots, na_n\}$ where $a_n \in \mathbb{N}$ (this follows from a uniform version of Theorem 5.1 that can be found in Bergelson et al. (2000)); every IP-set (Furstenberg and Katznelson 1985); the set $\{p(n), n \in \mathbb{N}\}$ where p is any nonconstant integer polynomial with $p(0) = 0$ (Bergelson and Leibman 1996) and, more generally, when the range of the polynomial contains multiples of an arbitrary integer (Frantzikinakis 2008); the set $\{p(n), n \in S\}$ where p is an integer polynomial with $p(0) = 0$ and S is any IP-set (Bergelson and McCutcheon 2000); and the set of values of an admissible generalized polynomial (Bergelson and McCutcheon 2010; McCutcheon 2005). Moreover, it was shown in Frantzikinakis et al. (2007) for $k = 2$ and in Wooley and Ziegler (2012) for general $k \in \mathbb{N}$ that the sets of shifted primes $\{p-1, p \text{ prime}\}$ (or the set $\{p+1, p \text{ prime}\}$) are sets of k -recurrence (see also Bergelson et al. (2011), Frantzikinakis et al. (2013), Koutsogiannis (2018a), and Sun (2015) for related work). Several other multiple recurrence results were obtained in the last 10 years, including results for Hardy sequences (Bergelson et al.

“Single and multiple recurrence...”; Frantzikinakis 2009, 2010; Frantzikinakis and Wierdl 2009), integer part polynomial sequences (Karageorgos and Koutsogiannis “Integer part independent...”; Koutsogiannis 2018a, b), random sequences (Frantzikinakis et al. 2012, 2016), and sets of arithmetic nature (Bergelson et al. “A structure theorem for...”; Frantzikinakis and Host 2017a, b).

More generally, one would like to know for which sequences of integers $a_1(n), \dots, a_k(n)$ it is the case that for every invertible measure-preserving system (X, B, μ, T) and $A \in B$ with $\mu(A) > 0$, there is some nonzero $n \in \mathbb{N}$ such that

$$\mu(A \cap T^{-a_1(n)}A \cap \dots \cap T^{-a_k(n)}A) > 0. \quad (9)$$

Unfortunately, a criterion analogous to the one given in Theorem 4.3 for 1-recurrence is not yet available for k -recurrence when $k > 1$. Nevertheless, there have been some notable positive results, such as the following:

Theorem 5.3 (Bergelson and Leibman 1996)

Let (X, B, μ, T) be an invertible measure-preserving system and $p_1(n), \dots, p_k(n)$ be integer polynomials with zero constant term. Then for every $A \in B$ with $\mu(A) > 0$, there is some $n \in \mathbb{N}$ such that

$$\mu(A \cap T^{-p_1(n)}A \cap \dots \cap T^{-p_k(n)}A) > 0. \quad (10)$$

Furthermore, it has been shown that the n in (10) can be chosen from any IP set (Bergelson and McCutcheon 2000) and the polynomials $p_1, \dots, p_k$ can be chosen to belong to the more general class of admissible generalized polynomials (McCutcheon 2005) or the class of intersective polynomials (Bergelson et al. 2008).

An important boost in the area of multiple recurrence was given by a breakthrough of Host and Kra (2005). Building on work of Conze and Lesigne (1984, 1988) and Furstenberg and Weiss (Furstenberg and Weiss 1996) (see also the excellent survey of Kra (2006a), exploring close

parallels with Green and Tao (2008), and the seminal paper of Gowers (2001)), they isolated the structured component (or factor) of a measure-preserving system that one needs to analyze in order to prove several multiple recurrence and convergence results. This allowed them, in particular, to prove existence of L^2 limits for the so-called “Furstenberg ergodic averages” $\frac{1}{N} \sum_{n=1}^N \prod_{i=0}^k f(T^{in}x)$, which had been a major open problem since the original ergodic proof of Szemerédi’s theorem. Subsequently Ziegler in Ziegler (2007) gave a new proof of the aforementioned limit theorem and established minimality of the factor in question. It turns out that this minimal component admits of a purely algebraic characterization; it is a *nilsystem*, i.e., a rotation on a homogeneous space of a nilpotent Lie group. This fact, coupled with some recent results about nilsystems (see Leibman (2005a, b) for example), makes the analysis of some otherwise intractable multiple recurrence problems much more manageable. These developments have made it possible to obtain new multiple recurrence results, and they also allowed us to estimate the size of the multiple intersection in (6) for $k = 2, 3$ (the case $k = 1$ is Theorem 3.1).

Theorem 5.4 (Bergelson et al. 2005) *Let (X, B, μ, T) be an ergodic measure-preserving system and $A \in B$. Then for $k = 2, 3$ and for every $\varepsilon > 0$,*

$$\begin{aligned} \mu(A \cap T^{-n}A \cap \dots \cap T^{-kn}A) \\ > \mu^{k+1}(A) - \varepsilon \end{aligned} \quad (11)$$

for a set of $n \in \mathbb{N}$ with bounded gaps.

Based on work of Ruzsa that appears as an appendix to the paper, it is also shown in Bergelson et al. (2005) that a similar estimate fails for ergodic systems (with any power of $\mu(A)$ on the right hand side) when $k \geq 4$. Moreover, when the system is nonergodic, it also fails for $k = 2, 3$, as can be seen with the help of an example in Behrend (1946). Again considering the doubling map $Tx = 2x$ and the set $A = [0, 1/2]$, one sees that the positive results for $k \leq 3$ are sharp.

When the polynomials $n, 2n, \dots, kn$ are replaced by linearly independent polynomials $p_1, p_2, \dots, p_k$ with zero constant term, similar lower bounds hold for every $k \in \mathbb{N}$ without assuming ergodicity (Frantzikinakis and Kra 2006). The case where the polynomials $n, 2n, 3n$ are replaced with general polynomials p_1, p_2, p_3 with zero constant term is treated in Frantzikinakis (2008) (see also Donoso et al. “Optimal lower bounds for multiple...”), and more general results involving Hardy field sequences and polynomials evaluated at the primes are obtained in Donoso et al. “Optimal lower bounds for multiple...”

Connections with Combinatorics and Number Theory

The combinatorial ramifications of ergodic-theoretic recurrence were first observed by Furstenberg, who perceived a correspondence between recurrence properties of measure-preserving systems and the existence of structures in sets of integers having positive upper density. This gave rise to the field of ergodic Ramsey theory, in which problems in combinatorial number theory are treated using techniques from ergodic theory. The following formulation is from Bergelson (1987b).

Theorem 6.1 *Let Λ be a subset of the integers. There exists an invertible measure-preserving system (X, B, μ, T) and a set $A \in B$ with $\mu(A) = \bar{d}(\Lambda)$ such that*

$$\begin{aligned} \bar{d}(\Lambda \cap (\Lambda - n_1) \cap \dots \cap (\Lambda - n_k)) \\ \geq \mu(A \cap T^{-n_1}A \cap \dots \cap T^{-n_k}A), \end{aligned} \quad (12)$$

for all $k \in \mathbb{N}$ and $n_1, \dots, n_k \in \mathbb{Z}$.

Proof The space X will be taken to be the sequence space $\{0, 1\}^{\mathbb{Z}}$, B is the Borel σ -algebra, while T is the shift map defined by $(Tx)(n) = x(n+1)$ for $x \in \{0, 1\}^{\mathbb{Z}}$, and A is the set of sequences x with $x(0) = 1$. So the only thing that depends on Λ is the measure μ which we now define. For $m \in \mathbb{N}$ set $\Lambda^0 = \mathbb{Z} \setminus \Lambda$ and $\Lambda^1 = \Lambda$. Using a diagonal argument,

we can find an increasing sequence of integers $(N_m)_{m \in \mathbb{N}}$ such that $\lim_{m \rightarrow \infty} |\Lambda \cap [1, N_m]|/N_m = \bar{d}$ (Λ) and such that

$$\lim_{m \rightarrow \infty} \frac{|(\Lambda^{i_1} - n_1) \cap (\Lambda^{i_2} - n_2) \cap \dots \cap (\Lambda^{i_r} - n_r) \cap [1, N_m]|}{N_m} \quad (13)$$

exists for every $n_1, \dots, n_r \in \mathbb{Z}$, and $i_1, \dots, i_r \in \{0, 1\}$. For $n_1, n_2, \dots, n_r \in \mathbb{Z}$, and $i_1, i_2, \dots, i_r \in \{0, 1\}$, we define the measure μ of the cylinder set $\{x_{n_1} = i_1, x_{n_2} = i_2, \dots, x_{n_r} = i_r\}$ to be the limit (13). Thus defined, μ extends to a pre-measure on the algebra of sets generated by cylinder sets and hence by Carathéodory's extension theorem (Carathéodory 1968) to a probability measure on B . It is easy to check that $\mu(A) = \bar{d}$ (Λ) , the shift transformation T preserves the measure μ and (12) holds.

Using this principle for $k = 1$, one may check that any set of recurrence is *intersective*, that is intersects $E - E$ for every set E of positive density. Using it for $n_1 = n, n_2 = 2n, \dots, n_k = kn$, together with Theorem 5.1, one gets an ergodic proof of Szemerédi's theorem (Szemerédi 1975), stating that every subset of the integers with positive upper density contains arbitrarily long arithmetic progressions (conversely, one can easily deduce Theorem 5.1 from Szemerédi's theorem and that intersective sets are sets of recurrence). Making the choice $n_1 = n^2$ and using part (iv) of Theorem 4.3, we get an ergodic proof of the surprising result of Sárközy (1978) stating that every subset of the integers with positive upper density contains two elements whose difference is a perfect square. More generally, using Theorem 6.1, one can translate all of the recurrence results of the previous two sections to results in combinatorics. (This is not straightforward for Theorem 5.4 because of the ergodicity assumption made there. We refer the reader to Bergelson et al. (2005) for the combinatorial consequence of this result.) We mention explicitly only the combinatorial consequence of Theorem 5.3.

Theorem 6.2 (Bergelson and Leibman 1996)

Let $\Lambda \subset \mathbb{Z}$ with $\bar{d}(\Lambda) > 0$ and $p_1, \dots, p_k$ be integer polynomials with zero constant term. Then Λ contains infinitely many configurations of the form $\{x, x + p_1(n), \dots, x + p_k(n)\}$.

The ergodic proof is the only one known for this result, although very recently some special cases were covered in Peluse and Prendiville ("Quantitative bounds in the...") and Prendiville (2017) using more elementary (but complicated) arguments.

Ergodic-theoretic contributions to the field of geometric Ramsey theory were made by Furstenberg et al. (1990), who showed that if E is a positive upper density subset of $\mathbb{R}^2$, then (i) E contains points with any large enough distance (see also Bourgain (1986) and Falconer and Marstrand (1986)) and (ii) every δ -neighborhood of E contains three points forming a triangle congruent to any given large enough dilation of a given triangle (in Bourgain (1986), it is shown that if the three points lie on a straight line, one cannot always find three points with this property in E itself). Moreover, a generalization of property (ii) to arbitrary finite configurations of $\mathbb{R}^m$ was obtained by Ziegler (2006).

We also mention some exciting connections of multiple recurrence with some structural properties of the set of prime numbers. The first one is in the work of Green and Tao (2008), where the existence of arbitrarily long arithmetic progressions of primes was demonstrated; the authors, in addition to using Szemerédi's theorem outright, use several ideas from its ergodic-theoretic proofs, as appearing in Furstenberg (1977) and Furstenberg et al. (1982). The second one is in the recent work of Tao and Ziegler (2008); a quantitative version of Theorem 5.3 was used to prove that the primes contain arbitrarily long polynomial progressions. Furthermore, results in ergodic theory related to the structure of the minimal characteristic factors of certain multiple ergodic averages play an important role in the work of Green, Tao, and Ziegler (Green and Tao 2010, 2012a, b, 2014; Green et al. 2012), where they get asymptotic formulas for the number of k -term arithmetic progressions of primes up to x . This work verifies an interesting special case of

the Hardy-Littlewood k -tuple conjecture predicting the asymptotic growth rate of $N_{a_1, \dots, a_k}(x)$ = the number of configurations of primes having the form $\{p, p + a_1, \dots, p + a_k\}$ with $p \leq x$.

In a more recent development, the tools developed in the last two decades to deal with delicate multiple recurrence problems have played an instrumental role in analyzing the structure of measure-preserving systems naturally associated with bounded multiplicative functions. These results were used in the last 2 years in works of Tao and Teräväinen ([The structure of logarithmically...; The structure of correlations...](#)) to make progress on the Chowla and Elliott conjectures and in works of Frantzikinakis and Host ([2018, Furstenberg systems of bounded...](#)) to make progress on the Möbius disjointness conjecture of Sarnak. It appears that this interplay of ergodic theory and number theory will be useful for the resolution of several notoriously difficult problems concerning higher-order correlations of multiplicative and other number theoretic functions.

Finally, we remark that in this article we have restricted attention to multiple recurrence and Furstenberg correspondence for $\mathbb{Z}$ -actions, while in fact there is a wealth of literature on extensions of these results to general commutative, amenable, and even non-amenable groups. For an excellent exposition of these and other recent developments, the reader is referred to the survey articles (Austin [“Multiple recurrence and finding...”](#); Bergelson [1996, 2006a, b](#)). Here, we give just one notable combinatorial corollary to some work of this kind, a density version of the classical Hales-Jewett coloring theorem (Hales and Jewett [1963](#)).

Theorem 6.3 (Furstenberg and Katznelson (1991); see also Polymath (2012)) *Let $W_n(A)$ denote the set of words of length n with letters in the alphabet $A = \{a_1, \dots, a_k\}$. For every $\varepsilon > 0$, there exists $N_0 = N_0(\varepsilon, k)$ such that if $n \geq N_0$, then any subset S of $W_n(A)$ with $|S| \geq \varepsilon k^n$ contains a combinatorial line, i.e., a set consisting of k n -letter words, having fixed letters in l positions, for some $0 \leq l < n$, the remaining $n - l$ positions*

being occupied by a variable letter x , for $x = a_1, \dots, a_k$. (For example, in $W_4(A)$, the sets $\{(a_1, x, a_2, x) : x \in A\}$ and $\{(x, x, x, x) : x \in A\}$ are combinatorial lines.)

At first glance, the uninitiated reader may not appreciate the importance of this “master” density result, so it is instructive to derive at least one of its immediate consequences. Let $A = \{0, 1, \dots, k - 1\}$ and interpret $W_n(A)$ as integers in base k having at most n digits. Then a combinatorial line in $W_n(A)$ is an arithmetic progression of length k – for example, the line $\{(a_1, x, a_2, x) : x \in A\}$ corresponds to the progression $\{m, m + n, m + 2n, m + 3n\}$, where $m = a_1 + a_2 d^2$ and $n = d + d^3$. This allows one to deduce Szemerédi’s theorem. Similarly, one can deduce from Theorem 6.3 multi-dimensional and IP extensions of Szemerédi’s theorem (Furstenberg and Katznelson [1979, 1985](#)) and some related results about vector spaces over finite fields (Furstenberg and Katznelson [1985](#)).

Future Directions

In this section we formulate a few open problems relating to the material in the previous three sections. It should be noted that this selection reflects the authors’ interests and does not strive for completeness. A more extensive list of problems related to ergodic theory of $\mathbb{Z}$ -actions can be found in Frantzikinakis ([2016](#)).

We start with an intriguing question of Katznelson ([2001](#)) about sets of topological recurrence. A set $S \subset \mathbb{N}$ is a set of Bohr recurrence if for every $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ and $\varepsilon > 0$ there exists $s \in S$ such that $\{s\alpha_i\} \in [0, \varepsilon] \cup [1 - \varepsilon, 1)$ for $i = 1, \dots, k$.

Problem 1 *Is every set of Bohr recurrence a set of topological recurrence?*

Background for this problem and evidence for a positive answer can be found in Katznelson ([2001](#)) and Weiss ([2000](#)). A negative answer for a related question concerning measure theoretic recurrence is given in Kriz ([1987](#)) (see also Griesmer [“Bohr topology and difference sets...”](#)).

Problem 2 Is the set $S = \{l2^m3^n : l, m, n \in \mathbb{N}\}$ a set of recurrence? Is it a set of k -recurrence for every $k \in \mathbb{N}$?

It can be shown that S is a set of Bohr recurrence. Theorem 4.3 cannot be applied since the uniform distribution condition fails for some irrational numbers α . A relevant question was asked by Bergelson in Bergelson (1996): “Is the set $S = \{2^m3^n : m, n \in \mathbb{N}\}$ good for single recurrence for weakly mixing systems?”

We mentioned in section “[Subsequence Recurrence](#)” that the sets of shifted primes and the set of fractional powers of integers are known to be sets of k -recurrence for every $k \in \mathbb{N}$. The following related question remains open.

Problem 3 Show that for every $c \in \mathbb{R}^+ \setminus \mathbb{Z}$, the set $\{[p^c], p \in \mathcal{P}\}$, where $\mathcal{P}$ is the set of primes, is a set of k -recurrence for every $k \in \mathbb{N}$.

The problem is open even for $k = 2$ and $c = \frac{3}{2}$.

We mentioned in section “[Subsequence Recurrence](#)” that random non-lacunary sequences (see definition there) are almost surely sets of recurrence.

Problem 4 Show that random non-lacunary sequences are almost surely sets of k -recurrence for every $k \in \mathbb{N}$.

Some progress is made in Bhattacharya et al. (“[Upper tails large deviations...](#)”), Briët et al. (2017), Briët and Gopi (“[Gaussian width bounds...](#)”), Christ (“[On random multilinear operator...](#)”), and Frantzikinakis et al. (2012, 2016), but the problem is open even for $k = 2$ for random sequences with at most quadratic growth. We refer the reader to the survey of Rosenblatt and Wierdl (1995) for a nice exposition of the argument used by Bourgain (Bourgain 1988) to handle the case $k = 1$.

It was shown in Frantzikinakis et al. (2006) that if S is a set of 2-recurrence, then the set of its squares is a set of recurrence for circle rotations. The same method shows that it is actually a set of Bohr recurrence.

Problem 5 If $S \subset \mathbb{Z}$ is a set of 2-recurrence, is it true that $S^2 = \{s^2 : s \in S\}$ is a set of recurrence?

A similar question was asked in Brown et al. (1999): “If S is a set of k -recurrence for every k , is the same true of S^2 ?”

One would like to find a criterion that would allow one to deduce that a sequence is good for double (or higher-order) recurrence from some uniform distribution properties of this sequence.

Problem 6 Find necessary conditions for double recurrence similar to the one given in Theorem 4.3.

It is now well understood that such a criterion should involve uniform distribution properties of some generalized polynomials or two-step nilsequences.

We mentioned in section “[Connections with Combinatorics and Number Theory](#)” that every positive upper density subset of $\mathbb{R}^2$ contains points with any large enough distance. Bourgain (1986) constructed a positive upper density subset E of $\mathbb{R}^2$, a triangle T , and numbers $t_n \rightarrow \infty$, such that E does not contain congruent copies of all t_n -dilations of T . But the triangle T used in this construction is degenerate, which leaves the following question open.

Problem 7 Is it true that every positive upper density subset of $\mathbb{R}^2$ contains a triangle congruent to any large enough dilation of a given non-degenerate triangle?

For further discussion on this question, the reader can consult the survey of Graham (1994).

The following question of Aaronson and Nakada (Aaronson 1981) is related to a classical question of Erdős concerning whether every $K \subset \mathbb{N}$ such that $\sum_{n \in K} 1/n = \infty$ contains arbitrarily long arithmetic progressions.

Problem 8 Suppose that $(X, \mathcal{B}, \mu, T)$ is a $\{1/n\}$ -conservative ergodic measure-preserving system. Is it true that for every $A \in \mathcal{B}$ with $\mu(A) > 0$ and $k \in \mathbb{N}$ we have $\mu(A \cap T^{-n}A \cap \dots \cap T^{-kn}A) > 0$ for some $n \in \mathbb{N}$?

The answer is positive for the class of Markov shifts, and it is remarked in Aaronson (1981) that if the Erdős conjecture is true, then the answer will be positive in general. The converse is not known to be true. For a related result showing that multiple recurrence is preserved by extensions of infinite measure-preserving systems, see Meyerovitch “On multiple and polynomial. . .”

Our next problem is motivated by the question whether Theorem 6.3 has a polynomial version (for a precise formulation of the general conjecture, see Bergelson (1996)). Not even this most special consequence of it is known to hold.

Problem 9 Let $\varepsilon > 0$. Does there exist $N = N(\varepsilon)$ having the property that every family P of subsets of $\{1, \dots, N\}^2$ satisfying $|P| \geq \varepsilon 2^{N^2}$ contains a configuration $\{A, A \cup (\gamma \times \gamma)\}$, where $A \subset \{1, \dots, N\}^2$ and $\gamma \subset \{1, \dots, N\}$ with $A \cap (\gamma \times \gamma) = \emptyset$?

A measure-preserving action of a general countably infinite group G is a function $g \rightarrow T_g$ from G into the space of measure-preserving transformations of a probability space X such that $T_{gh} = T_g T_h$. It is easy to show that a version of Khintchine’s recurrence theorem holds for such actions: if $\mu(A) > 0$ and $\varepsilon > 0$, then $\{g : \mu(A \cap T_g A) > (\mu(A)^2 - \varepsilon)\}$ is syndetic. However it is unknown whether the following ergodic version of Roth’s theorem holds.

Problem 10 Let (T_g) and (S_g) be measure-preserving G -actions of a probability space X that commute in the sense that $T_g S_h = S_h T_g$ for all $g, h \in G$. Is it true that for all positive measure sets A , the set of g such that $\mu(A \cap T_g A \cap S_g A) > 0$ is syndetic?

We remark that for general (possibly amenable) groups G not containing arbitrarily large finite subgroups nor elements of infinite order, it is not known whether one can find a *single* such $g \neq e$. On the other hand, the answer is known to be positive for general G in case $(T_g^{-1} S_g)$ is a G -action (Bergelson and McCutcheon 2007); even under such strictures, however, it is unknown whether a triple recurrence theorem holds.

References

- Aaronson J (1981) The asymptotic distribution behavior of transformations preserving infinite measures. *J Anal Math* 39:203–234
- Aaronson J (1997) An introduction to infinite ergodic theory. Mathematical surveys and monographs, vol 50. American Mathematical Society, Providence
- Austin T. Multiple recurrence and finding patterns in dense sets. In: Badziahin D, Gorodnik A, Peyerimhoff N (eds) Dynamics and analytic number theory. London mathematical society lecture note series. Cambridge University Press, Cambridge, pp 189–257
- Barreira L (2001) Hausdorff dimension of measures via Poincaré recurrence. *Commun Math Phys* 219:443–463
- Barreira L (2005) Poincaré recurrence: old and new. In: XIVth international congress on mathematical physics. World Scientific Publishing, Hackensack, pp 415–422
- Behrend F (1946) On sets of integers which contain no three in arithmetic progression. *Proc Natl Acad Sci U S A* 23:331–332
- Bergelson V (1987a) Weakly mixing PET. *Ergodic Theory Dyn Syst* 7:337–349
- Bergelson V (1987b) Ergodic Ramsey theory (ed: Simpson S, Logic and combinatorics). *Contemp Math* 65:63–87
- Bergelson V (1996) Ergodic Ramsey theory – an update, Ergodic Theory of $\mathbb{Z}^d$ -actions (ed: Pollicott M, Schmidt K). London Math Soc Lecture Note Ser 228:1–61
- Bergelson V (2000) The multifarious Poincaré recurrence theorem. In: Foreman M, Kechris A, Louveau A, Weiss B (eds) Descriptive set theory and dynamical systems. London mathematical society lecture note series, vol 277. Cambridge University Press, Cambridge, pp 31–57
- Bergelson V (2006a) Combinatorial and Diophantine applications of Ergodic theory. Appendix A by A. Leibman and Appendix B by A. Quas and M. Wierdl. In: Hasselblatt B, Katok A (eds) Handbook of dynamical systems, vol. 1B. Elsevier, Amsterdam, pp 745–841
- Bergelson V (2006b) Ergodic Ramsey theory: a dynamical approach to static theorems. In: Proceedings of the international congress of mathematicians, vol II. Madrid, pp 1655–1678
- Bergelson V, Leibman A (1996) Polynomial extensions of van der Waerden’s and Szemerédi’s theorems. *J Am Math Soc* 9:725–753
- Bergelson V, McCutcheon R (2000) An ergodic IP polynomial Szemerédi theorem. *Mem Am Math Soc* 146: viii+106 pp
- Bergelson V, McCutcheon R (2007) Central sets and a noncommutative Roth theorem. *Am J Math* 129:1251–1275
- Bergelson V, McCutcheon R (2010) Idempotent ultrafilters, multiple weak mixing and Szemerédi’s theorem for generalized polynomials. *J Anal Math* 111:77–130
- Bergelson V, Furstenberg H, McCutcheon R (1996) IP-sets and polynomial recurrence. *Ergodic Theory Dyn Syst* 16:963–974

- Bergelson V, Host B, McCutcheon R, Parreau F (2000) Aspects of uniformity in recurrence. *Colloq Math* 84/85(Part 2):549–576
- Bergelson V, Host B, Kra B, with an appendix by Ruzsa I (2005) Multiple recurrence and nilsequences. *Invent Math* 160(2):261–303
- Bergelson V, Håland-Knutson I, McCutcheon R (2006) IP systems, generalized polynomials and recurrence. *Ergodic Theory Dyn Syst* 26:999–1019
- Bergelson V, Leibman A, Lesigne E (2008) Intersective polynomials and the polynomial Szemerédi theorem. *Adv Math* 219(1):369–388
- Bergelson V, Leibman A, Ziegler T (2011) The shifted primes and the multidimensional Szemerédi and polynomial van der Waerden Theorems. *Comptes Rendus Mathématique* 349(3–4):123–125
- Bergelson V, Kolesnik G, Son Y (2019) Uniform distribution of subpolynomial functions along primes and applications. *J Anal Math* 137(1):135–187
- Bergelson V, Kułaga-Przymus J, Lemańczyk M. A structure theorem for level sets of multiplicative functions and applications. To appear in *Int Math Res Not*. arXiv:1708.02613
- Bergelson V, Moreira J, Richter FK. Single and multiple recurrence along non-polynomial sequences. Preprint, arXiv:1711.05729
- Bhattacharya B, Ganguly S, Shao X, Zhao Y. Upper tails large deviations for arithmetic progressions in a random set. To appear in *Int Math Res Note*. arXiv:1605.02994
- Birkhoff G (1931) A proof of the ergodic theorem. *Proc Natl Acad Sci U S A* 17:656–660
- Boshernitzan M (1993) Quantitative recurrence results. *Invent Math* 113:617–631
- Boshernitzan M, Kolesnik G, Quas A, Wierdl M (2005) Ergodic averaging sequences. *J Anal Math* 95:63–103
- Bourgain J (1986) A Szemerédi type theorem for sets of positive density in $\mathbb{R}^k$. *Israel J Math* 54(3):307–316
- Bourgain J (1988) On the maximal ergodic theorem for certain subsets of the positive integers. *Israel J Math* 61:39–72
- Briët J, Gopi S. Gaussian width bounds with applications to arithmetic progressions in random settings. To appear in *Int Math Res Note*. arXiv:1711.05624
- Briët J, Dvir Z, Gopi S (2017) Outlaw distributions and locally decodable codes. *Proc ITCS* arXiv:1609.06355
- Brown T, Graham R, Landman B (1999) On the set of common differences in van der Waerden’s theorem on arithmetic progressions. *Can Math Bull* 42:25–36
- Carathéodory C (1968) *Vorlesungen über reelle Funktionen*, 3rd edn. Chelsea Publishing, New York
- Chazottes J, Ugalde E (2005) Entropy estimation and fluctuations of hitting and recurrence times for Gibbsian sources. *Discrete Contin Dyn Syst Ser B* 5(3):565–586
- Christ M. On random multilinear operator inequalities. Unpublished manuscript. Available at arXiv:1108.5655
- Conze J, Lesigne E (1984) Théorèmes ergodiques pour des mesures diagonales. *Bull Soc Math France* 112(2):143–175
- Conze J, Lesigne E (1988) Sur un théorème ergodique pour des mesures diagonales. In: *Probabilités*, Publ Inst Rech Math Rennes, 1987-1. University of Rennes I, Rennes, pp 1–31
- Donoso S, Le AN, Moreira J, Sun W. Optimal lower bounds for multiple recurrence. To appear in *Ergodic Theory Dyn Syst*. arXiv:1809.06912
- Einsiedler M, Ward T (2011) *Ergodic theory with a view towards number theory*. Graduate texts in mathematics, vol 259. Springer London, London
- Evans D, Searles D (2002) The fluctuation theorem. *Adv Phys* 51:1529–1585
- Falconer K, Marstrand J (1986) Plane sets with positive density at infinity contain all large distances. *Bull Lond Math Soc* 18:471–474
- Forrest A (1991) The construction of a set of recurrence which is not a set of strong recurrence. *Israel J Math* 76:215–228
- Frantzikinakis N (2008) Multiple ergodic averages for three polynomials and applications. *Trans Am Math Soc* 360(10):5435–5475
- Frantzikinakis N (2009) Equidistribution of sparse sequences on nilmanifolds. *J Anal Math* 109:353–395
- Frantzikinakis N (2010) Multiple recurrence and convergence for Hardy sequences of polynomial growth. *J Anal Math* 112:79–135
- Frantzikinakis N (2016) Some open problems on multiple ergodic averages. *Bull Hell Math Soc* 60:41–90
- Frantzikinakis N, Host B (2017a) Higher order Fourier analysis of multiplicative functions and applications. *J Am Math Soc* 30:67–157
- Frantzikinakis N, Host B (2017b) Multiple ergodic theorems for arithmetic sets. *Trans Am Math Soc* 369(10):7085–7105
- Frantzikinakis N, Host B (2018) The logarithmic Sarnak conjecture for ergodic weights. *Ann Math* 187:869–931
- Frantzikinakis N, Host B. Furstenberg systems of bounded multiplicative functions and applications. To appear in *Int Math Res Note IMRN*. arXiv:1804.08556
- Frantzikinakis N, Kra B (2006) Ergodic averages for independent polynomials and applications. *J London Math Soc* 74(1):131–142
- Frantzikinakis N, Wierdl M (2009) A Hardy field extension of Szemerédi’s theorem. *Adv Math* 222:1–43
- Frantzikinakis N, Lesigne E, Wierdl M (2006) Sets of k -recurrence but not $(k+1)$ -recurrence. *Ann Inst Fourier* 56(4):839–849
- Frantzikinakis N, Host B, Kra B (2007) Multiple recurrence and convergence for sets related to the primes. *J Reine Angew Math* 611:131–144
- Frantzikinakis N, Lesigne E, Wierdl M (2012) Random sequences and pointwise convergence of multiple ergodic averages. *Indiana Univ Math J* 61:585–617
- Frantzikinakis N, Host B, Kra B (2013) The polynomial multidimensional Szemerédi Theorem along shifted primes. *Israel J Math* 194:331–348
- Frantzikinakis N, Lesigne E, Wierdl M (2016) Random differences in Szemerédi’s theorem and related results. *J Anal Math* 130:91–133
- Furstenberg H (1977) Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J Anal Math* 71:204–256

- Furstenberg H (1981) Recurrence in ergodic theory and combinatorial number theory. Princeton University Press, Princeton
- Furstenberg H, Katznelson Y (1979) An ergodic Szemerédi theorem for commuting transformations. *J Anal Math* 34:275–291
- Furstenberg H, Katznelson Y (1985) An ergodic Szemerédi theorem for IP-systems and combinatorial theory. *J Anal Math* 45:117–168
- Furstenberg H, Katznelson Y (1991) A density version of the Hales-Jewett theorem. *J Anal Math* 57:64–119
- Furstenberg H, Weiss B (1996) A mean ergodic theorem for $(1/N)\sum_{n=1}^N f(T^n x)g(T^{n^2} x)$. Convergence in ergodic theory and probability (Columbus, OH, 1993), Ohio State Univ Math Res Inst Publ, vol 5, de Gruyter, Berlin, pp 193–227
- Furstenberg H, Katznelson Y, Ornstein D (1982) The ergodic theoretical proof of Szemerédi’s theorem. *Bull Am Math Soc* 7(3):527–552
- Furstenberg H, Katznelson Y, Weiss B (1990) Ergodic theory and configurations in sets of positive density. In: *Mathematics of Ramsey theory. Algorithms combinatorics*, vol 5. Springer, Berlin, pp 184–198
- Galatolo S, Kim DH, Park KK (2006) The recurrence time for ergodic systems with infinite invariant measures. *Nonlinearity* 19:2567–2580
- Glasner E (2003) Ergodic theory via joinings. *Mathematical surveys and monographs*, vol 101. American Mathematical Society, Providence
- Gowers W (2001) A new proof of Szemerédi’s theorem. *Geom Funct Anal* 11:465–588
- Graham RL (1994) Recent trends in Euclidean Ramsey theory. *Trends in discrete mathematics. Discrete Math* 136(1–3):119–127
- Green B, Tao T (2008) The primes contain arbitrarily long arithmetic progressions. *Ann Math* 167:481–547
- Green B, Tao T (2010) Linear equations in primes. *Ann Math* 171:1753–1850
- Green B, Tao T (2012a) The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann Math* 175:465–540
- Green B, Tao T (2012b) The Möbius function is strongly orthogonal to nilsequences. *Ann Math* 175:541–566
- Green B, Tao T (2014) On the quantitative distribution of polynomial nilsequences- erratum. *Ann Math* 179:1175–1183, arXiv:1311.6170v3
- Green B, Tao T, Ziegler T (2012) An inverse theorem for the Gowers $U^{s+1}[N]$ -norm. *Ann Math* 176(2):1231–1372
- Griesmer J. Bohr topology and difference sets for some abelian groups. Preprint, arXiv:1608.01014
- Hales A, Jewett R (1963) Regularity and positional games. *Trans Am Math Soc* 106:222–229
- Hasley T, Jensen M (2004) Hurricanes and butterflies. *Nature* 428:127–128
- Host B, Kra B (2005) Nonconventional ergodic averages and nilmanifolds. *Ann Math* 161:397–488
- Host B, Kra B (2018) Nilpotent structures in Ergodic theory. *Mathematical surveys and monographs*, vol 236. American Mathematical Society, Providence
- Kac M (1947) On the notion of recurrence in discrete stochastic processes. *Bull Am Math Soc* 53:1002–10010
- Kamae T, Mendés-France M (1978) Van der Corput’s difference theorem. *Israel J Math* 31:335–342
- Karageorgos D, Koutsogiannis A. Integer part independent polynomial averages and applications along primes. To appear in *Studia Math.* arXiv:1708.06820
- Katznelson Y (2001) Chromatic numbers of Cayley graphs on $\mathbb{Z}$ and recurrence. Paul Erdős and his mathematics (Budapest, 1999). *Combinatorica* 21(2):211–219
- Khinchine A (1934) Eine Verschärfung des Poincaréschen “Wiederkehrrsatzes”. *Comp Math* 1:177–179
- Koutsogiannis A (2018a) Closest integer polynomial multiple recurrence along shifted primes. *Ergodic Theory Dyn Syst* 38:666–685
- Koutsogiannis A (2018b) Integer part polynomial correlation sequences. *Ergodic Theory Dyn Syst* 38:1525–1542
- Kra B (2006a) The Green-Tao theorem on arithmetic progressions in the primes: an ergodic point of view. *Bull Am Math Soc* 43:3–23
- Kra B (2006b) From combinatorics to ergodic theory and back again. In: *Proceedings of international congress of mathematicians*, vol III. Madrid, pp 57–76
- Kra B (2007) Ergodic methods in additive combinatorics. In: *Additive combinatorics, CRM proceedings and lecture notes*, vol 43. American Mathematical Society, Providence, pp 103–143
- Kra B (2011) Poincare recurrence and number theory: thirty years later. *Bull Am Math Soc* 48:497–501
- Kriz I (1987) Large independent sets in shift invariant graphs. Solution of Bergelson’s problem. *Graphs Comb* 3:145–158
- Leibman A (2002) Lower bounds for ergodic averages. *Ergodic Theory Dyn Syst* 22:863–872
- Leibman A (2005a) Pointwise convergence of ergodic averages for polynomial sequences of rotations of a nilmanifold. *Ergodic Theory Dyn Syst* 25:201–213
- Leibman A (2005b) Pointwise convergence of ergodic averages for polynomial actions of $\mathbb{Z}^d$ by translations on a nilmanifold. *Ergodic Theory Dyn Syst* 25:215–225
- McCutcheon R (1995) Three results in recurrence. In: *Ergodic theory and its connections with harmonic analysis* (Alexandria, 1993). London mathematical society lecture note series, vol 205. Cambridge University Press, Cambridge, pp 349–358
- McCutcheon R (2005) FVIP systems and multiple recurrence. *Israel J Math* 146:157–188
- Meyerovitch T. On multiple and polynomial recurrent extensions of infinite measure preserving transformations. Unpublished. Available at arXiv:0703914v2
- Ornstein D, Weiss B (1993) Entropy and data compression schemes. *IEEE Trans Inform Theory* 39:78–83
- Peluse S, Prendiville S. Quantitative bounds in the nonlinear Roth theorem. Preprint, arXiv:1903.02592
- Petersen K (1989) Ergodic theory. *Cambridge studies in advanced mathematics*, vol 2. Cambridge University Press, Cambridge

- Poincaré H (1890) Sur le problème des trois corps et les équations de la dynamique. *Acta Math* 13:1–270
- Polymath DHJ (2012) A new proof of the density Hales-Jewett theorem. *Ann Math* 175:1283–1327
- Prendiville S (2017) Quantitative bounds in the polynomial Szemerédi theorem: the homogeneous case. *Discrete Anal* 5:1–34
- Rosenblatt J, Wierdl M (1995) Pointwise ergodic theorems via harmonic analysis. In: *Ergodic theory and its connections with harmonic analysis* (Alexandria, 1993). London mathematical society lecture note series, vol 205. Cambridge University Press, Cambridge, pp 3–151
- Sárközy A (1978) On difference sets of integers III. *Acta Math Acad Sci Hungar* 31:125–149
- Shkredov I (2002) Recurrence in the mean. *Mat Zametki* 72(4):625–632; translation in *Math Notes* (2002) 72(3–4):576–582
- Sklar L (2004) Philosophy of statistical mechanics. In: Zalta EN (ed) *The Stanford encyclopedia of philosophy* (Summer 2004 Edition). <http://plato.stanford.edu/archives/sum2004/entries/statphys-statmech/>
- Sun W (2015) Multiple recurrence and convergence for certain averages along shifted primes. *Ergodic Theory Dyn Syst* 35(5):1592–1609
- Szemerédi E (1975) On sets of integers containing no k elements in arithmetic progression. *Acta Arith* 27:299–345
- Tao T, Teräväinen J. The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures. To appear in *Duke Math J*. arXiv:1708.02610
- Tao T, Teräväinen J. The structure of correlations of multiplicative functions at almost all scales, with applications to the Chowla and Elliott conjectures. Preprint, arXiv:1809.02518
- Tao T, Ziegler T (2008) The primes contain arbitrarily long polynomial progressions. *Acta Math* 201:213–305
- von Neumann J (1932) Proof of the Quasi-ergodic hypothesis. *Proc Natl Acad Sci U S A* 18(1):70–82
- Walters P (1982) *An introduction to ergodic theory*. Graduate texts in mathematics, vol 79. Springer, New York/Berlin
- Weiss B (2000) *Single orbit dynamics*. CBMS regional conference series in mathematics, vol 95. American Mathematical Society, Providence
- Wooley T, Ziegler T (2012) Multiple recurrence and convergence along the primes. *Am J Math* 134:1705–1732
- Wyner A, Ziv J (1989) Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Trans Inform Theory* 35:1250–1258
- Zermelo E (1896) Über einen Satz der Dynamik und die mechanische Wärmetheorie. *Ann Phys* 57:485–494; English translation (1966) On a theorem of dynamics and the mechanical theory of heat. In: Brush SG (ed) *Kinetic theory*, vol II. Oxford, pp 208–217
- Ziegler T (2006) Nilfactors of $\mathbb{R}^m$ -actions and configurations in sets of positive upper density in $\mathbb{R}^m$. *J Anal Math* 99:249–266
- Ziegler T (2007) Universal characteristic factors and Furstenberg averages. *J Am Math Soc* 20:53–97

Ergodic Theorems

Andrés del Junco

Department of Mathematics, University of
Toronto, Toronto, ON, Canada

Article Outline

Glossary

Definition of the Subject

Introduction

Ergodic Theorems for Measure-Preserving Maps

Generalizations to Continuous Time and Higher-
Dimensional Time

Pointwise Ergodic Theorems for Operators

Subadditive and Multiplicative Ergodic Theorems

Entropy and the Shannon–McMillan–Breiman
Theorem

Amenable Groups

Subsequence and Weighted Theorems

Ergodic Theorems and Multiple Recurrence

Rates of Convergence

Ergodic Theorems for Non-amenable Groups

Future Directions

Bibliography

Glossary

Automorphism a dynamical system $T: X \rightarrow X$, where X is a measure space and T is an invertible map preserving measure.

Dynamical system in its broadest sense, any set X , with a map $T: X \rightarrow X$. The classical example is: X is a set whose points are the states of some physical system and the state x is succeeded by the state Tx after one unit of time.

Ergodic average if f is a function on X let $A_n f(x) = n^{-1} \sum_{i=0}^{n-1} f(T^i x)$; the average of the values of f over the first n points in the orbit of x .

Ergodic theorem an assertion that ergodic averages converge in some sense.

Iteration repeated applications of the map T above to arrive at the state of the system after n units of time.

Maximal inequality an inequality which allows one to bound the pointwise oscillation of a sequence of functions. An essential tool for proving pointwise ergodic theorems.

Mean ergodic theorem an assertion that ergodic averages converge with respect to some norm on a space of functions.

Operator any linear operator U on a vector space of functions on X , for example one arising from a dynamical system T by setting $Uf(x) = f(Tx)$. More generally any linear transformation on a real or complex vector space.

Orbit of x the forward images $x, Tx, T^2x, \dots$ of $x \in X$ under iteration of T . When T is invertible one may consider the forward, backward or two-sided orbit of x .

Pointwise ergodic theorem an assertion that ergodic averages $A_n f(x)$ converge for some or all $x \in X$, usually for a.e. x .

Positive contraction an operator T on a space of functions endowed with a norm $\|\cdot\|$ such that T maps positive functions to positive functions and $\|Tf\| \leq \|f\|$.

Stationary process a sequence $(X_1, X_2, \dots)$ of random variables (real or complex-valued measurable functions) on a probability space whose joint distributions are invariant under shifting $(X_1, X_2, \dots)$ to $(X_2, X_3, \dots)$.

Uniform distribution a sequence $\{x_n\}$ in $[0, 1]$ is uniformly distributed if for each interval $I \subset [0, 1]$, the time it spends in I is asymptotically proportional to the length of I .

Definition of the Subject

Ergodic theorems are assertions about the long-term statistical behavior of a dynamical system. The subject arose out of Boltzmann's ergodic hypothesis which sought to equate the spatial average of a function over the set of states in a physical system having a fixed energy with the

time average of the function observed by starting with a particular state and following its evolution over a longtime period.

Introduction

Suppose that $(X, \mathcal{B}, \mu)$ is a measure space and $T : (X, \mathcal{B}, \mu) \rightarrow (X, \mathcal{B}, \mu)$ is a measurable and measure-preserving transformation, that is $\mu(T^{-1}E) = \mu(E)$ for all $E \in \mathcal{B}$. One important motivation for studying such maps is that a Hamiltonian physical system (see the article by Petersen in this collection) gives rise to a one-parameter group $\{T_t : t \in \mathbb{R}\}$ of maps in the phase space of the system which preserve Lebesgue measure. The ergodic theorem of Birkhoff asserts that for $f \in L_1(\mu)$

$$\frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \quad (1)$$

converges a.e. and that if T is *ergodic* (to be defined shortly) then the limit is $\int f d\mu$. This may be viewed as a justification for Boltzmann's ergodic hypothesis that "space averages equal time averages". See Zund (2002) for some history of the ergodic hypothesis. For physicists, then, the problem is reduced to showing that a given physical system is ergodic, which can be very difficult. However ergodic systems arise in many natural ways in mathematics. One example is rotation $z \mapsto \lambda z$ of the unit circle if λ is a complex number of modulus one which is not a root of unity. Another is the shift transformation on a sequence of i.i.d. random variables, for example a coin-tossing sequence. Another is an automorphism of a compact Abelian group. Often a transformation possesses an invariant measure which is not obvious at first sight. Knowledge of such a measure can be a very useful tool. See Petersen's article for more examples.

If $(X, \mathcal{B}, \mu)$ is a probability space and T is ergodic then Birkhoff's ergodic theorem implies that if A is a measurable subset of X then for almost every x , the frequency with which x visits

A is asymptotically equal to $\mu(A)$, a very satisfying justification of intuition. For example, applying this to the coin-tossing sequence one obtains the *strong law of large numbers* which asserts that almost every infinite sequence of coin tosses has tails occurring with asymptotic frequency $\frac{1}{2}$. One also obtains Borel's theorem on normal numbers which asserts that for almost all $x \in [0, 1]$ each digit $0, 1, 2, \dots, 9$ occurs with limiting frequency $\frac{1}{10}$. The so-called *continued fraction transformation* $x \mapsto x^{-1} \bmod 1$ on $(0, 1)$ has a finite invariant measure $\frac{dx}{1+x}$. (Throughout this article $x \bmod 1$ denotes the fractional part of x .) Applying Birkhoff's theorem then gives precise information about the frequency of occurrence of any $n \in \mathbb{N}$ in the continued fraction expansion of x , for a.e. x . See for example Billingsley (Billingsley 1965).

These are the classical roots of the subject of ergodic theorems. The subject has evolved from these simple origins into a vast field in its own right, quite independent of physics or probability theory. Nonetheless it still has close ties to both these areas and has also forged new links with many other areas of mathematics.

Our purpose here is to give a broad overview of the subject in a historical perspective. There are several excellent references, notably the books of Krengel (1985) and Tempelman (1992) which give a good picture of the state of the subject at the time they appeared. There has been tremendous progress since then. The time is ripe for a much more comprehensive survey of the field than is possible here.

Many topics are necessarily absent and many are only glimpsed. For example this article will not touch on random ergodic theorems. See the articles (Durand and Schneider 2003; Lemańczyk et al. n.d.) for some references on this topic.

I thank Mustafa Akcoglu, Ulrich Krengel, Michael Lin, Dan Rudolph and particularly Joe Rosenblatt and Vitaly Bergelson for many helpful comments and suggestions. I would like to dedicate this article to Mustafa Akcoglu who has been such an important contributor to the development of ergodic theorems over the past 40 years. He has

also played a vital role in my mathematical development as well as of many other mathematicians. He remains a source of inspiration to me, and a valued friend.

Ergodic Theorems for Measure-Preserving Maps

Suppose $(X, \mathcal{B}, \mu)$ is a measure space. A (measure-preserving) *endomorphism* of $(X, \mathcal{B}, \mu)$ is a measurable mapping $T : X \rightarrow X$ such that $\mu(T^{-1}E) = \mu(E)$ for any measurable subset $E \subset X$. If T has a measurable inverse then one says that it is an *automorphism*. In this article the unqualified terms “endomorphism” or “automorphism” will always mean measure-preserving. T is called *ergodic* if for all measurable E one has

$$T^{-1}E = E \Rightarrow \mu(E) = 0 \quad \text{or} \quad \mu(E^c) = 0.$$

A very general class of examples comes from the notion of a *stationary stochastic process* in probability theory. A stochastic process is a sequence of measurable functions $f_1, f_2, \dots$ on a probability space $(X, \mathcal{B}, \mu)$ taking values in a measurable space $(Y, \mathcal{C})$. The *distribution* of the process, a measure ν on $Y^{\mathbb{N}}$, is defined as the image of μ under the map $(f_1, f_2, \dots) : X \rightarrow Y^{\mathbb{N}}$. ν captures all the essential information about the process $\{f_i\}$. In effect one may view any process $\{f_i\}$ as a probability measure on $Y^{\mathbb{N}}$. The process is said to be *stationary* if ν is invariant under the left shift transformation S on $Y^{\mathbb{N}}$, $S(y)(i) = y(i+1)$, that is S is an endomorphism of the probability space $(Y^{\mathbb{N}}, \nu)$.

From a probabilistic point of view the most natural examples of stationary stochastic processes are independent identically distributed processes, namely the case when ν is a product measure $\lambda^{\mathbb{N}}$ for some measure λ on $(Y, \mathcal{C})$. More generally one can consider a stationary Markov process defined by transition probabilities on the state space Y and an invariant probability on Y . See for example Chap. 7 in (Breiman 1968), also section “Pointwise Ergodic Theorems for Operators” below.

The first and most fundamental result about endomorphisms is the celebrated recurrence theorem of Poincaré (1987).

Theorem 1 *Suppose μ is finite, $A \in \mathcal{B}$ and $\mu(A) > 0$. Then for a.e. $x \in A$ there is an $n > 0$ such that $T^n x \in A$, in fact there are infinitely many such n .*

It may be viewed as an ergodic theorem, in that it is a qualitative statement about how x behaves under iteration of T . For a proof, observe that if $E \subset A$ is the measurable set of points which never return to A then for each $n > 0$ the set $T^{-n}E$ is disjoint from E . Applying T^{-m} one gets also $T^{-(n+m)}E \cap T^{-m}E = \emptyset$. Thus $E, T^{-1}E, T^{-2}E, \dots$ is an infinite sequence of disjoint sets all having measure $\mu(E)$ so $\mu(E) = 0$ since $\mu(X) < \infty$. Much later Kac (1947) formulated the following quantitative version of Poincaré’s theorem.

Theorem 2 *Suppose that $\mu(X) = 1$, T is ergodic and let $r_A x$ denote the time of first return to A , that is $r_A(x)$ is the least $n > 0$ such that $T^n x \in A$. Then*

$$\frac{1}{\mu(A)} \int_A r_A d\mu = \frac{1}{\mu(A)}, \quad (2)$$

that is, the expected value of the return time to A is $\mu(A)^{-1}$.

Koopman (1931) made the observation that associated to an automorphism T there is a unitary operator $U = U_T$ defined on the Hilbert space $L_2(\mu)$ by the formula $Uf = f \circ T$. This led von Neumann (von Neumann 1932) to prove his mean ergodic theorem.

Theorem 3 *Suppose $\mathcal{H}$ is a Hilbert space, U is a unitary operator on $\mathcal{H}$ and let P denote the orthogonal projection on the subspace of U -invariant vectors. Then for any $x \in \mathcal{H}$ one has*

$$\left\| \frac{1}{n} \sum_{i=0}^{n-1} U^i x - Px \right\| \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (3)$$

Von Neumann's theorem is usually quoted as above but to be historically accurate he dealt with unitary operators indexed by a continuous time parameter.

Inspired by von Neumann's theorem Birkhoff very soon proved his pointwise ergodic theorem (Birkhoff 1931). In spite of the later publication by von Neumann his result did come first. See (Bergelson 2007; Zund 2002) for an interesting discussion of the history of the two theorems and of the interaction between Birkhoff and von Neumann.

Theorem 4 *Suppose $(X, \mathcal{B}, \mu)$ is a measure space and T is an endomorphism of $(X, \mathcal{B}, \mu)$. Then for any $f \in L_1 = L_1(\mu)$ there is a T -invariant function $g \in L_1$ such that*

$$A_n f(x) = \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \rightarrow g(x) \quad \text{a.e.} \quad (4)$$

Moreover if μ is finite then the convergence also holds with respect to the L_1 norm and one has $\int_E g \, d\mu = \int_E f \, d\mu$ for all T -invariant subsets E .

Again this formulation of Birkhoff's theorem is not historically accurate as he dealt with a smooth flow on a manifold. It was soon observed that the theorem, and its proof, remain valid for an abstract automorphism of a measure space although the realization that T need not be invertible seems to have taken a little longer.

The notation $A_n f = A_n(T)f$ as above will occur often in the sequel. Whenever T is an endomorphism one uses the notation $Tf = f \circ T$ and with this notation $A_n(T) = \frac{1}{n} \sum_{i=0}^{n-1} T^i$. When the scalars ($\mathbb{R}$ or $\mathbb{C}$) for an L_1 space are not specified the notation should be understood as referring to either possibility. In most of the theorems in this article the complex case follows easily from the real and any indications about proofs will refer to the real case.

Although von Neumann originally used spectral theory to prove his result, there is a quick proof, attributed to Riesz by Hopf in his 1937 book (Hopf 1937), which uses only elementary properties of Hilbert space. Let I denote the

(closed) subspace of U -invariant vectors and I' the (usually not closed) subspace of vectors of the form $f - Uf$. It is easy to check that any vector orthogonal to I' must be in I , whence the subspace $I + I'$ is dense in $\mathcal{H}$. For any vector of the form $x = y + y'$, $y \in I$, $y' \in I'$ it is clear that $A_n x = \frac{1}{n} \sum_{i=0}^{n-1} U^i x$ converges to $y = Px$, since $A_n y = y$ and if $y' = z - Uz$ then the telescoping sum $A_n y' = n^{-1}(z - U^n z)$ converges to 0. This establishes the desired convergence for $x \in I + I'$ and it is easy to extend it to the closure of $I + I'$ since the operators A_n are contractions of $\mathcal{H}$ ($\|A_n\| \leq 1$). Lorch (1939) used a soft argument in a similar spirit to extend von Neumann's theorem from the case of a unitary operator on a Hilbert space to that of an arbitrary linear contraction on any reflexive Banachspace. Sine (1970) gave a necessary and sufficient condition for the strong convergence of the ergodic averages of a contraction on an arbitrary Banach space.

Birkhoff's theorem has the distinction of being one of the most reproved theorems of twentieth century mathematics. One approach to the pointwise convergence, parallel to the argument just seen, is to find a dense subspace E of L_1 so that the convergence holds for all $f \in E$ and then try to extend the convergence to all $f \in L_1$ by an approximation argument. The first step is not too hard. For simplicity assume that μ is a finite measure. As in the proof of von Neumann's theorem the subspace E spanned by the T -invariant L_1 functions together with functions of the form $g - Tg$, $g \in L_1$, is dense in $L_1(\mu)$. This can be seen by using the Hahn–Banach theorem and the duality of L_1 and L_∞ . (Here one needs finiteness of μ to know that $L_\infty \subset L_1$.) The pointwise convergence of $A_n f$ for invariant f is trivial and for $f = g - Tg$ it follows from telescoping of the sum and the fact that $n^{-1} T^n g \rightarrow 0$ a.e. This last can be shown by using the Borel–Cantelli lemma. The second step, extending pointwise convergence, as opposed to norm convergence, for f in a dense subspace to all f in L_1 is a delicate matter, requiring a *maximal inequality*.

Roughly speaking a maximal inequality is an inequality which bounds the pointwise oscillation of $A_n f$ in terms of the norm of f . The now standard

maximal inequality in the context of Birkhoff's theorem is the following, due to Kakutani and Yosida (1939). Birkhoff's proof of his theorem includes a weaker version of this result. Let $S_n f = n A_n f$, the n th partial sum of the iterates of f .

Theorem 5 *Given any real $f \in L_1$ let $A = \cup_{n \geq 1} \{S_n f \geq 0\}$. Then*

$$\int_A f \, d\mu \geq 0. \quad (5)$$

Moreover if one sets $Mf = \sup_{n \geq 1} A_n f$ then for any $\alpha > 0$

$$\mu\{Mf > \alpha\} \leq \frac{1}{\alpha} \|f\|_1 \quad (6)$$

A distributional inequality such as (6) will be referred to as a *weak L_1 inequality*. Note that (6) follows easily from (5) by applying (5) to $f - \alpha$, at least in the case when μ is finite. With the maximal inequality in hand it is straightforward to complete the proof of Birkhoff's theorem. For a real-valued function f let

$$\text{Osc } f = \limsup A_n f - \liminf A_n f. \quad (7)$$

$\text{Osc } f = 0$ a.e. if and only if $A_n f$ converges a.e. (to a possibly infinite limit). One has $\limsup A_n f \leq Mf \leq M|f|$ and by symmetry $\liminf A_n f \geq M|f|$, so $\text{Osc } f \leq 2M|f|$. To establish the convergence of $A_n f$ for a real-valued $f \in L_1$ let $\epsilon > 0$ and write $f = g + h$ with $g \in E$ (the subspace where convergence has already been established), $h \in L^1$ and $\|h\|_1 < \epsilon$. Then since $\text{Osc } g = 0$ one has $\text{Osc } f = \text{Osc } h$. Thus for any fixed $\alpha > 0$, using (6)

$$\begin{aligned} \mu\{\text{Osc } f > \alpha\} &= \mu\{\text{Osc } h > \alpha\} \leq \mu\left\{Mh > \frac{\alpha}{2}\right\} \\ &\leq \frac{2\|h\|_1}{\alpha} < \frac{2\epsilon}{\alpha}. \end{aligned} \quad (8)$$

Since $\epsilon > 0$ was arbitrary one concludes that $\mu\{\text{Osc } f > \alpha\} = 0$ and since $\alpha > 0$ is arbitrary it follows that $\mu\{\text{Osc } f > 0\} = 0$, establishing the

a.e. convergence. Moreover a simple application of Fatou's lemma shows that the limiting function is in L_1 , hence finite a.e.

There are many proofs of (5). Two of particular interest are Garsia's (1970), perhaps the shortest and most mysterious, and the proof via the filling scheme of Chacón and Ornstein (1960), perhaps the most intuitive, which goes like this. Given a function $g \in L_1$ write $g^+ = \max(g, 0)$, $g^- = g^+ - g$ and let $Ug = Tg^+ - g^-$. Interpretation: the region between the graph of g^+ and the X -axis is a sheaf of vertical spaghetti sticks, the intervals $[0, g^+(x)]$, $x \in X$, and $-g^-$ is a hole. Now move the spaghetti (horizontally) by T and then let it drop (vertically) into the hole leaving a new hole and a new sheaf which are the negative and positive parts of Ug . Now let $E' = \cup_{n \geq 1} \{U^n f \geq 0\}$, the set of points x at which the hole is eventually filled after finitely many iterations of U .

The key point is that $E = E'$. Indeed if $S_n f(x) \geq 0$ for some n , then the total linear height of sticks over $x, Tx, \dots, T^{n-1}x$ is greater than the total linear depth of holes at these points. The only way that spaghetti can escape from these points is by first filling the hole at x , which shows $x \in E'$. Similar thinking shows that if $x \in E'$ and the hole at x is filled for the first time at time n then $S_n f(x) \geq 0$, so $x \in E$, and that all the spaghetti that goes into the hole at x comes from points $T^i x$ which belong to E' . This shows that $E = E'$ and that the part of the hole lying beneath E is eventually filled by spaghetti coming from $E' = E$. Thus the amount of spaghetti over E is no less than the size of the hole under E , that is $\int_E f \, d\mu \geq 0$.

Most proofs of Birkhoff's theorem use a maximal inequality in some form but a few avoid it altogether, for example (Katok and Hasselblatt 1995; Shields 1987). It is also straightforward to deduce Birkhoff's theorem directly from a maximal inequality, as Birkhoff does, without first establishing convergence on a dense subspace. However the technique of proving a pointwise convergence theorem by finding an appropriate dense subspace and a suitable maximal inequality has proved extremely useful, not only in ergodic theory.

Indeed, in some sense maximal inequalities are unavoidable: this is the content of the following principle proved already in 1926 by Banach. The principle has many formulations; the following one is a slight simplification of the one to be found in (Krengel 1985). Suppose B is a Banach space, $(X, \mathcal{B}, \mu)$ is a finite measure space and let E denote the space of μ -equivalence classes of measurable real-valued functions on X . A linear map $T: B \rightarrow E$ is said to be *continuous in measure* if for each $\epsilon > 0$

$$\|x_n - x\| \rightarrow 0 \Rightarrow \mu\{|Tx_n - Tx| > \epsilon\} \rightarrow 0.$$

Suppose that T_n is a sequence of linear maps from B to E which are continuous in measure and let $Mx = \sup_n |T_n x|$. Of course if $T_n x$ converges a.e. to a finite limit then $Mx < \infty$ a.e.

Theorem 6 (Banach Principle) *Suppose $Mx < \infty$ a.e. for each $x \in X$. Then there is a function $C(\lambda)$ such that $C(\lambda) \rightarrow 0$ as $\lambda \rightarrow \infty$ such that for all $x \in B$ and $\lambda > 0$ one has*

$$\mu\{Mx \geq \lambda \|x\|\} \leq C(\lambda). \quad (9)$$

Moreover the set of x for which $T_n x$ converges a.e. is closed in B .

The first chapter of Garsia's book (1970) contains a nice introduction to the Banach principle.

It should be noted that, for general integrable f , Mf need not be in L^1 . However if $f \in L^p$ for $p > 1$ then Mf does belong to L^p and one has the estimate

$$\|Mf\|_p \leq \frac{p}{p-1} \|f\|_p. \quad (10)$$

This can be derived from (6) using also the (obvious) fact that $\|Mf\|_\infty \leq \|f\|_\infty$. Any such estimate on the norm of a maximal function will be called a *strong L_p* inequality. It also follows from (6) that if $\mu(X)$ is finite and $f \in L \log L$, that is $\int |f| \log^+ |f| d\mu < \infty$, then $Mf \in L^1$. In fact Ornstein (1971) has shown that the converse of this last statement holds provided T is ergodic.

There is a special setting where one has uniform convergence in the ergodic theorem. Suppose T is a homeomorphism of a compact metric space X . By a theorem of Krylov and Bogoliouboff (1937) there is at least one probability measure on the Borel σ -algebra of X which is invariant under T . T is said to be *uniquely ergodic* if there is only one Borel probability measure, say μ , invariant under T . It is easy to see that when this is the case then T is an ergodic automorphism of $(X, \mathcal{B}, \mu)$. As an example, if α is an irrational number then the rotation $z \mapsto e^{2\pi i \alpha} z$ is a uniquely ergodic transformation of the circle $\{z : |z| = 1\}$. Equivalently $x \mapsto x + \alpha \pmod{1}$ is a uniquely ergodic map on $[0, 1]$. A quick way to see this is to show that the Fourier coefficients $\hat{\mu}(n)$ of any invariant probability μ are zero for $n \neq 0$. The Jewett–Krieger theorem (see Jewett (1969) and Krieger (1972)) guarantees that unique ergodicity is ubiquitous in the sense that any automorphism of a probability space is measure-theoretically isomorphic to a uniquely ergodic homeomorphism. The following important result is due to Oxtoby (1952)).

Theorem 7 *If T is uniquely ergodic, μ is its unique invariant probability measure and $f \in C(X)$ then the ergodic averages $A_n(f)$ converge uniformly to $\int f d\mu$.*

This result can be proved along the same lines as the proof given above of von Neumann's theorem. In a nutshell, one uses the fact that the dual of $C_{\mathbb{R}}(X)$ is the space of finite signed measures on X and the unique ergodicity to show that functions of the form $f - f \circ T$ together with the invariant functions (which are just constant functions) span a dense subspace of $C(X)$.

A sequence $\{x_n\}$ in the interval $[0, 1]$ is said to be *uniformly distributed* if $\frac{1}{n} \sum_{i=1}^n f(x_i) \rightarrow \int_0^1 f(x) dx$ for any $f \in C[0, 1]$ (equivalently for any Riemann integrable f or for any $f = 1_I$ where I is any subinterval of $[0, 1]$). As a simple application of Theorem 7 one obtains the linear case of the following result of Weyl (1916).

Theorem 8 *If α is irrational and $p(x)$ is any non-constant polynomial with integer coefficients then the sequence $\{p(n)\alpha \bmod 1\}$ is uniformly distributed.*

Furstenberg (1967), see also (Furstenberg 1981), has shown that Weyl's result, in full generality, can be deduced from the unique ergodicity of certain affine transformations of higher dimensional tori. For polynomials of degree $k > 1$ Weyl's result is usually proved by inductively reducing to the case $k = 1$, using the following important lemma of van der Corput (see for example (Kuipers and Niederreiter 1974)).

Theorem 9 *Suppose that for each fixed $h > 0$ the sequence*

$$\{x_{n+h} - x_n \bmod 1\}_n$$

is uniformly distributed. Then $\{x_n\}$ is uniformly distributed.

When μ is infinite and T is ergodic the limiting function in Birkhoff's theorem is 0 a.e. In 1937 Hopf (1937) proved a generalization of Birkhoff's theorem which is more meaningful in the case of an infinite invariant measure. It is a special case of a later theorem of Hurewicz (1944), which we will discuss first.

Suppose that $(X, \mathcal{B}, \mu)$ is a σ -finite measure space. If ν is another σ -finite measure on $\mathcal{B}$ write $\nu \ll \mu$ (ν is *absolutely continuous* relative to μ) if $\mu(E) = 0$ implies $\nu(E) = 0$ and one writes $\nu \sim \mu$ if $\nu \ll \mu$ and $\mu \ll \nu$. Consider a *non-singular automorphism* $\tau : X \rightarrow X$, meaning that τ is measurable with a measurable inverse and that $\mu(E) = 0$ if and only if $\mu(\tau E) = 0$. In other words $\nu = \mu \circ \tau \sim \mu$. By the Radon–Nikodym theorem there is a function $\rho \in L_1(\mu)$ such that $\rho > 0$ a.e. and $\nu(E) = \int_E \rho \, d\mu$ for all measurable E . In order to obtain an associated operator T on L_1 which is an (invertible) isometry one defines

$$Tf(x) = \rho(x)f(\tau x). \quad (11)$$

The dual operator on L_∞ is then given by $T^*g = g \circ \tau^{-1}$.

If ν is a σ -finite measure equivalent to μ which is invariant under τ then the ergodic theory of τ can be reduced to the measure-preserving case using σ . The interesting case is when there is no such ν . It was an open problem for some time whether there is always an equivalent invariant measure. In 1960 Ornstein (1960) gave an example of a τ which does not have an equivalent invariant measure. It is curious that, with hindsight, such examples were already known in the fifties to people studying von Neumann algebras.

For $f \in L_1$ let $S_n f = \sum_{i=0}^n T^i f$. τ is said to be *conservative* if there is no set E with $\mu(E) > 0$ such that $\tau^{-i}E$, $i = 0, 1, \dots$ are pairwise disjoint, that is if the Poincaré recurrence theorem remains valid. For example, the shift on $\mathbb{Z}$ is not conservative. Hurewicz (1944) proved the following ratio ergodic theorem.

Theorem 10 *Suppose τ is conservative, $f, g \in L_1$ and $g(x) > 0$ a.e. Then $S_n f / S_n g$ converges a.e. to a τ -invariant limit h . If τ is ergodic then $h = \frac{\int f \, d\mu}{\int g \, d\mu}$.*

In the case when μ is τ -invariant one has $Tf = f \circ \tau$. If μ is invariant and finite, taking $g = 1$ one recovers Birkhoff's theorem. If μ is invariant and σ -finite then Hurewicz's theorem becomes the theorem of Hopf alluded to earlier.

Wiener and Wintner (1941) proved the following variant of Birkhoff's theorem.

Theorem 11 (Wiener–Wintner) *Suppose T is an automorphism of a probability space $(X, \mathcal{B}, \mu)$. Then for any $f \in L_1$ there is a subset $X' \subset X$ of measure one such that for each $x \in X'$ and $\lambda \in \mathbb{C}$ of modulus 1 the sequence $\frac{1}{n} \sum_{i=0}^{n-1} \lambda^i T^i f(x)$ converges.*

It is an easy consequence of the ergodic theorem that one has a.e. convergence for a given f and λ but the point here is that the set on which the convergence occurs is independent of λ .

Generalizations to Continuous Time and Higher-Dimensional Time

A (measure-preserving) flow is a one-parameter group $\{T_t, t \in \mathbb{R}\}$ of automorphisms of $(X, \mathcal{B}, \mu)$, that is $T_{t+s} = T_t T_s$, such that $T_t x$ is measurable as a function of (t, x) . It will always be implicitly assumed that the map $(t, x) \mapsto T_t x$ from $\mathbb{R} \times X$ to X is measurable. Theorem 4 generalizes to flows by replacing sums with integrals and this generalization follows without difficulty from Theorem 4. (As already observed this observation reverses the historical record.) Theorem 4 may be viewed as a theorem about the “discrete flow” $\{T_n = T^n : n \in \mathbb{Z}\}$. Wiener was the first to generalize Birkhoff’s theorem to families of automorphisms $\{T_g\}$ indexed by groups more general than $\mathbb{R}$ or $\mathbb{Z}$.

A measure-preserving flow is an *action* of $\mathbb{R}$ while a single automorphism corresponds to an action of $\mathbb{Z}$. A (measure-preserving) action of a group G is a homomorphism $T : g \mapsto T_g$ from G into the group of automorphisms of a measure space $(X, \mathcal{B}, \mu)$ (satisfying the appropriate joint measurability condition in case G is not discrete). Suppose now that $G = \mathbb{R}^k$ or $\mathbb{Z}^k$ and T is an action of G on $(X, \mathcal{B}, \mu)$. In the case of $\mathbb{Z}^k$ an action amounts to an arbitrary choice of commuting maps $T_1, \dots, T_k$, $T_i = T(e_i)$ where e_i is the standard basis of $\mathbb{Z}^k$. In the case of $\mathbb{R}^k$ one must specify k commuting flows.

Let m denote counting measure on G in case $G = \mathbb{Z}^k$ and Lebesgue measure in case $G = \mathbb{R}^k$. For any subset E of G with $m(E) < \infty$ let

$$A_E f(x) = \frac{1}{m(E)} \int_E f(T_g x) dm(g). \quad (12)$$

One may then ask whether $A_E f$ converges to a limit, either in the mean or pointwise, as E varies through some sequence of sets which “grow large” or, in case $G = \mathbb{R}^k$, “shrink to 0”. The second case is referred to as a *local ergodic theorem*. In the case of ergodic theorems at infinity the continuous and discrete theories are rather similar and often the continuous analogue of a discrete result can be deduced from the discrete result.

In Wiener (1939) proved the following result for actions of $G = \mathbb{R}^k$ and ergodic averages over Euclidean balls $B_r = \{x \in \mathbb{R}^k : \|x\|_2 \leq r\}$.

Theorem 12 *Suppose T is an action of $\mathbb{R}^k$ on $(X, \mathcal{B}, \mu)$ and $f \in L_1(\mu)$.*

- (a) *For $f \in L_1$ $\lim_{r \rightarrow \infty} A_{B_r} f = g$ exists a.e. If μ is finite the convergence also holds with respect to the L_1 -norm, g is T -invariant and $\int_I g d\mu = \int_I f d\mu$ for every T -invariant set I .*
- (b) *$\lim_{r \rightarrow 0} A_{B_r} f = f$ a.e.*

The local aspect of Wiener’s theorem is closely related to the Lebesgue differentiation theorem, see for example Proposition 3.5.4 of (Krantz and Parks 1999), which, in its simplest form, states that for $f \in L^1(\mathbb{R}^k, m)$ one has a.e. convergence of $\frac{1}{m(B_r)} \int_{B_r} f(x+t) dt$ to $f(x)$ as $r \rightarrow 0$. The local ergodic theorem implies Lebesgue’s theorem, simply by considering the action of $\mathbb{R}^k$ on itself by translation. In fact the local ergodic theorem can also be deduced from Lebesgue’s theorem by a simple application of Fubini’s theorem (see for example (Krengel 1985), Chap. 1, Theorem 2.4 in the case $k = 1$).

The key point in Wiener’s proof is the following weak L_1 maximal inequality, similar to (6).

Theorem 13 *Let $Mf = \sup_{r>0} |A_{B_r} f|$. Then one has*

$$\mu\{Mf > \alpha\} \leq \frac{C}{\alpha} \|f\|_1, \quad (13)$$

where C is a constant depending only on the dimension d . (In fact one may take $C = 3^d$.)

In the case when T is the action of $\mathbb{R}^k$ on itself by translation (13) is the well-known maximal inequality for the Hardy–Littlewood maximal function (Krantz and Parks 1999, Lemma 3.5.3). Wiener proves (13) by way of the following *covering lemma*. If B is a ball in $\mathbb{R}^k$ let B' denote the concentric ball with three times the radius.

Theorem 14 Suppose a compact subset K of $\mathbb{R}^k$ is covered by a (finite) collection $\mathcal{U}$ of open balls. Then there exist pairwise disjoint $B_i \in \mathcal{U}$, $i = 1, \dots, k$ such that $\cup_i B'_i$ covers K .

To find the B_i it suffices to let B_1 be the largest ball in $\mathcal{U}$, then B_2 the largest ball which does not intersect B_1 and in general B_n the largest ball which does not intersect $\cup_{i=1}^{n-1} B_i$. Then it is not hard to argue that the B'_i cover K .

In general, a covering lemma is, roughly speaking, an assertion that, given a cover $\mathcal{U}$ of a set K in some space (often a group), one may find a subcollection $\mathcal{U}' \subset \mathcal{U}$ which covers a substantial part of K and is in some sense efficient in that $\sum_{U \in \mathcal{U}'} 1_U \leq C$, C some absolute constant. Covering lemmas play an important role in the proofs of many maximal inequalities. See Sect. 3.5 of (Krantz and Parks 1999) for a discussion of several of the best-known classical covering lemmas.

As Wiener was likely aware, the same kind of covering argument easily leads to maximal inequalities and ergodic theorems for averages over “sufficiently regular” sets. For example, in the case of $G = \mathbb{Z}^k$ use the standard total order $<$ on G and for $n \in \mathbb{N}^k$ let $S_n = \{m \in \mathbb{Z}^k : 0 < m < n\}$. Let $e(n)$ denote the maximum value of n_i/n_j . Then one has the following result.

Theorem 15 For any $e > 0$ and any integrable f

$$\lim_{n \rightarrow \infty, e(n) < e} A_{S_n} f$$

exists a.e. and the limit is characterized by the same properties as in Theorem 12. Moreover there is a maximal inequality

$$\mu \left\{ \sup_{e(n) < e} |A_{S_n} f| > \alpha \right\} \leq \frac{C}{\alpha} \|f\|_1, \quad (14)$$

where C is a constant depending only on e and the dimension k .

If one does not impose a bound on $e(n)$ the a.e. convergence of $A_{S_n} f$ may fail for $f \in L_1$, (although it does hold for any increasing sequence

of n 's by Theorem 30 below). Nonetheless Dunford (1951) and independently Zygmund (1951) showed that one does have unrestricted convergence if $f \in L_p$ for some $p > 1$, provided μ is finite. Let $T_1, \dots, T_k$ be any k (possibly non-commuting!) automorphisms of a finite measure space $(X, \mathcal{B}, \mu)$. For $n = (n_1, n_2, \dots, n_k)$ let $T_n = T_1^{n_1} \dots T_k^{n_k}$. (This is not an action of $\mathbb{Z}^k$ unless the T_i commute with each other.) As before when $F \subset \mathbb{Z}^k$ is a finite subset write $A_F f = \frac{1}{|F|} \sum_{n \in F} T_n f$. Finally let $P_j f = \lim_n A_n(T_j) f$.

Theorem 16 For $f \in L_p$

$$\lim_{n \rightarrow \infty} A_{S_n} f = P_1 \dots P_k f \quad (15)$$

both a.e. and in L_p .

The proof of Theorem 16 uses repeated applications of Birkhoff's theorem and hinges on (10).

Somewhat surprisingly Hurewicz's theorem was not generalized to higher dimensions until the very recent work of Feldman (2007). In fact the theorem fails if one considers averages over the cube $[0, n-1]^d$ in $\mathbb{Z}^d$. However Feldman was able to prove a suitably formulated generalization of Hurewicz's theorem for averages over symmetric cubes.

For $f \in L^1(\mathbb{R})$ the classical Hilbert transform is defined for a.e. t by

$$Hf(t) = \frac{1}{\pi} \lim_{\epsilon \rightarrow 0^+} \int_{|s| > \epsilon} \frac{f(t-s)}{s} ds \quad (16)$$

$$\text{Let } H^*f(t) = \frac{1}{\pi} \sup_{\epsilon > 0} \left| \int_{|s| > \epsilon} \frac{f(t-s)}{s} ds \right|, \quad \text{the}$$

corresponding maximal function. The proof that the limit (16) exists a.e. is based on the maximal inequality

$$\mu\{Hf > \lambda\} \leq \frac{C}{\lambda} \|f\|_1, \quad (17)$$

where C is an absolute constant. See Loomis (1946) for a proof of (17) using real-variable methods. Cotlar (1955b) proved the existence of the following ergodic analogue of the Hilbert

transform (actually for the n -dimensional Hilbert transform).

Theorem 17 *Suppose $\{T_t\}$ is a measure-preserving flow on a probability space $(X, \mathcal{B}, \mu)$ and $f \in L^1$. Then*

$$\lim_{\epsilon \rightarrow 0} \int_{\epsilon < t < \epsilon^{-1}} \frac{f(T_t x)}{t} dt \quad (18)$$

exists for a.a. x .

Motivated by Cotlar's result Calderón (1968) proved a general *transfer principle* which allows one to transfer maximal inequalities and convergence theorems for functions of a real or integer variable to the ergodic setting. Although stated for functions on $\mathbb{R}$ it applies equally well to $\mathbb{R}^k$ or $\mathbb{Z}^k$. This principle subsumes Birkhoff's theorem, Wiener's theorem and Cotlar's result. For simplicity only a special case of Calderón's result, in the discrete case, will be stated here.

Let T be an automorphism of a probability space $(X, \mathcal{B}, \mu)$, let m denote counting measure on $\mathbb{Z}$ and suppose σ is a probability measure on $\mathbb{Z}$. For $g \in L_1(m)$ define $\sigma(g)(n) = \sum_{i \in \mathbb{Z}} \sigma(i)g(n+i)$. For $f \in L_1(\mu)$ let $\sigma(T)f = \sum_{i \in \mathbb{Z}} \sigma(i)T^i f$, which is easily seen to converge a.e. and in L_1 -norm. Given a fixed sequence σ_n of probabilities define $Mg = \sup_n \sigma_n g$ and $M(T)f = \sup_n \sigma_n(T)f$.

Theorem 18 (Calderón) *Suppose there is a constant C such that*

$$m\{Mg > \lambda\} < \frac{C}{\lambda} \|g\|_1 \quad \text{for all } g \in L_1(m). \quad (19)$$

Then one also has

$$\mu\{M(T)f > \lambda\} < \frac{C}{\lambda} \|f\|_1 \quad \text{for all } T \quad \text{and} \\ g \in L_1(\mu). \quad (20)$$

In other words, in order to prove a maximal inequality for general T it suffices to prove it in

case T is the shift map on the integers. The idea of transference could already be seen in Wiener's proof of the $\mathbb{R}^n$ ergodic theorem. Transfer principles in various forms have become an important tool in the study of ergodic theorems. See Bellow (1999) for a very readable overview.

Pointwise Ergodic Theorems for Operators

Early in the history of ergodic theorems there were attempts to generalize the ergodic theorem to more general linear operators on L_p spaces, that is, operators which do not arise by composition with a mapping of X . In the case $p = 1$ the main motivation for this comes from the theory of Markov processes.

If $(X, \mathcal{B})$ is a measurable space a *sub-stochastic kernel* on X is a non-negative function P on $X \times \mathcal{B}$ such that

- (a) for each $x \in X$ $P(x, \cdot) = P_x$ is a measure on $\mathcal{B}$ such that $P_x(X) \leq 1$ and,
- (b) $P(\cdot, A)$ is a measurable function for each $A \in \mathcal{B}$.

It is most intuitive to think about the *stochastic* case, namely when each P_x is a probability measure. One then views $P(x, A)$ as the probability that the point x moves into the set A in one unit of time, so one has stochastic dynamics as opposed to deterministic dynamics, namely the case when $P_x = \delta_{Tx}$ for a map T . In this case the measures P_x are called *transition probabilities*.

If μ is a σ -finite measure on X one may define the measure $P\mu = \int P_x d\mu(x)$. $P\nu$ is also meaningful if ν is a finite signed measure. The case when $P\mu = \mu$ is the stochastic analogue of measure-preserving dynamics and the case when $P\mu \ll \mu$ is the analogue of non-singular dynamics. It is easy to see that given any σ -finite measure λ there is always a μ such that $\lambda \ll \mu$ and $P\mu \ll \mu$. Let $\tilde{L}_1$ denote the space of finite signed measures ν such that $\nu \ll \mu$, which is identified with $L_1 = L_1(\mu, \mathbb{R})$ via the Radon-Nikodym theorem. If $P\mu \ll \mu$ then P maps $\tilde{L}_1(\mu)$ into itself so the

restriction of P is an operator T on $L_1(\mu)$. T is a positive contraction, that is $\|T\| \leq 1$ and T maps non-negative functions to non-negative functions. As proved in, for example, (Neveu 1965), every positive contraction arises in this way from a substochastic kernel under the assumption that $(X, \mathcal{B})$ is standard. This simply means that there is some complete metric on X for which $\mathcal{B}$ is the σ -algebra of Borel sets. Virtually all measurable spaces encountered in analysis are standard so this should be viewed as a technicality only. See (Foguel 1969; Krengel 1985; Neveu 1965) for more about the relation between kernels and positive contractions.

The case when X is finite, P is stochastic and μ is a probability measure is classical in probability theory. P and μ determine a probability measure ν on $X^{\mathbb{N}}$ characterized by its values on cylinder sets, namely for all $x_1, \dots, x_n \in X$

$$\nu\{x \in X^{\mathbb{N}} : x(i) = x_i, 1 \leq i \leq n-1\} = \mu(x_1) \prod_{i=1}^{n-1} P_{x_i}(x_{i+1}). \quad (21)$$

The co-ordinate functions $X_i(x) = x(i)$ on the space $X^{\mathbb{N}}$ endowed with the probability ν form a Markov process, which will be stationary if and only if μ is P -invariant. For a general X the analogous construction is possible provided X is standard.

Hopf (1954) initiated the systematic study of positive L_1 contractions and proved the following ergodic theorem.

Theorem 19 *Suppose $(X, \mathcal{B}, \mu)$ is a probability space and T is a positive contraction on $L^1(\mu)$ satisfying $T1 = 1$ and $T^*1 = 1$. Then $\lim_{n \rightarrow \infty} n^{-1} \sum_{k=0}^{n-1} T^k f$ exists a.e.*

The importance of Hopf's article lies less in this convergence result than in the methods he developed. He proved that the maximal inequality (5) generalizes to all positive L_1 contractions T and used this to obtain the decomposition of X into its conservative and dissipative parts C and D , characterized by the fact that for any

$p \in L_1$ such that $p > 0$ a.e. $\sum_{k=0}^{\infty} T^k p$ is infinite a.e. on C and finite a.e. on D . These results are the cornerstone of much of the subsequent work on L_1 contractions.

Theorem 19 contains Birkhoff's theorem for an automorphism τ of $(X, \mathcal{B}, \mu)$, simply by defining $Tf = f \circ \tau$. In fact one can also deduce Theorem 19 from Birkhoff's theorem (if one assumes only that X is standard). Indeed the hypotheses of Hopf's theorem imply that the kernel P associated to T is stochastic ($T^*1 = 1$) and that μ is P -invariant ($T1 = 1$). Hopf's theorem then follows by applying Birkhoff's theorem to the shift on the stationary Markov process associated to P and μ . In fact Kakutani (1940) (see also Doob (1938)) had already made essentially the same observation, except that his result assumes the stationary Markov process to be already given.

In 1955 Dunford and Schwartz (1955) made essential use of Hopf's work to prove the following result.

Theorem 20 *Suppose $\mu(X) = 1$ and T is a (not necessarily positive) contraction with respect to both the L_1 and L_{∞} norms. Then the conclusion of Hopf's theorem remains valid.*

Note that the assumption that T contracts the L_{∞} -norm is meaningful, as $L_{\infty} \subset L_1$. The proof of the result is reduced to the positive case by defining a positive contraction $|T|$ analogous to the total variation of a complex measure.

Then in 1960 Chacón and Ornstein (1960) proved a definitive ratio ergodic theorem for all positive contractions of L_1 which generalizes both Hopf's theorem and the Hurewicz theorem.

Theorem 21 *Suppose T is a positive contraction of $L_1(\mu)$, where μ is σ -finite, $f, g \in L_1$ and $g \geq 0$. Then*

$$\frac{\sum_{i=0}^n T^i f}{\sum_{i=0}^n T^i g} \quad (22)$$

converges a.e. on the set $\{\sum_{i=0}^{\infty} T^i g > 0\}$.

In 1963 Chacón (1963a) proved a very general theorem for non-positive operators which includes the Chacón–Ornstein theorem as well as the Dunford–Schwartz theorem.

Theorem 22 *Suppose T is a contraction of L_1 and $p_n \geq 0$ is a sequence of measurable functions with the property that*

$$g \in L_1, \quad |g| \leq p_n \Rightarrow |Tg| \leq p_{n+1}. \quad (23)$$

Then

$$\frac{\sum_{i=0}^n T^i f}{\sum_{i=0}^n p_i} \quad (24)$$

converges a.e. to a finite limit on the set $\{\sum_{i=0}^{\infty} p_i > 0\}$.

If T is an L_1, L_{∞} -contraction and $p_n = 1$ for all n then the hypotheses of this theorem are satisfied, so Theorem 22 reduces to the result of Dunford and Schwartz. See (Chacón 1963b) for a concise overview of all of the above theorems in this section and the relations between them.

The identification of the limit in the Chacón–Ornstein theorem on the conservative part C of X is a difficult problem. It was solved by Neveu (1961) in case $C = X$, and in general by Chacón (1962). Chacón (1964) has shown that there is a non-singular automorphism τ of $(X, \mathcal{B}, \mu)$ such that for the associated invertible isometry T of $L_1(\mu)$ given by (11) there is an $f \in L_1(\mu)$ such that $\limsup A_n f = \infty$ and $\liminf A_n f = -\infty$.

In 1975 Akcoglu (1975) solved a major open problem when he proved the following celebrated theorem.

Theorem 23 *Suppose $T : L_p \mapsto L_p$ is a positive contraction. Then $A_n f = \frac{1}{n} \sum_{i=0}^{n-1} T^i f$ converges a.e. Moreover one has the strong L_p inequality*

$$\left\| \sup_n |A_n f| \right\|_p \leq \frac{p}{p-1} \|f\|_p. \quad (25)$$

As usual, the maximal inequality (25) is the key to the convergence. Note that it is identical in

form to (10) the classical strong L_p inequality for automorphisms. (25) was proved by A. Ionesco–Tulcea (now Bellow) (Tulcea 1964) in the case of positive invertible isometries of L_p . It is a result of Banach (1993), see also (Lamperti 1958), that in this case T arises from a non-singular automorphism τ of $(X, \mathcal{B}, \mu)$ in the form $Tf = \rho^{1/p} f \circ \tau$. By a series of reductions Bellow was able to show that in this case (25) can be deduced from (10).

Akcoglu's brilliant idea was to consider a *dilation* S of T which is a positive invertible isometry on a larger L_p space $\tilde{L}_p = L_p(Y, \mathcal{C}, \nu)$. What this means is that there is a positive isometric injection $D : L_p \rightarrow \tilde{L}_p$ and a positive projection P on $\tilde{L}_p$ whose range is $D(L_p)$ such that $DT^n = PS^n D$ for all $n \geq 0$. Given the existence of such an S it is not hard to deduce (25) for T from (25) for S . In fact Akcoglu constructs a dilation only in the case when L_p is finite dimensional and shows how to reduce the proof of (25) to this case. In the finite dimensional case the construction is very concrete and P is a conditional expectation operator. Later Akcoglu and Kopp (1977) gave a construction in the general case. It is noteworthy that the proof of Akcoglu's theorem consists ultimately of a long string of reductions to the classical strong L_p inequality (10), which in turn is a consequence of (5).

Subadditive and Multiplicative Ergodic Theorems

Consider a family $\{X_{n,m}\}$ of real-valued random variables on a probability space indexed by the set of pairs $(n, m) \in \mathbb{Z}^2$ such that $0 \leq n < m$. $\{X_{(n,m)}\}$ is called a (stationary) *subadditive process* if

- (a) the joint distribution of $\{X_{n,m}\}$ is the same as that of $\{X_{n+1, m+1}\}$,
- (b) $X_{n,m} \leq X_{n,l} + X_{l,m}$ whenever $n < l < m$.

Denoting the index set by $\{n < m\} \subset \mathbb{Z}^2$ the distribution of the process is a measure μ on $\mathbb{R}^{\{n < m\}}$ which is invariant under the shift Tx ($n, m) = x(n+1, m+1)$. Thus there is no loss of generality in assuming that there is an

underlying endomorphism T such that $X_{n,m} \circ T = X_{n+1,m+1}$. In 1968 Kingman (1968) proved the following generalization of Birkhoff's theorem. $\gamma = \inf \frac{1}{n} \int X_{0,n} d\mu$ is called the *time constant* of the process.

Theorem 24 *If the $X_{n,m}$ are integrable and $\gamma > -\infty$ then $\frac{1}{n}X_{0,n}$ converges a.e. and in L_1 -norm to a $\mathbb{T}$ -invariant limit $\bar{X} \in L_1(\mu)$ satisfying $\int \bar{X} d\mu = \gamma$.*

It is easy to deduce from the above that if one assumes only that $X_{0,1}^+$ is integrable then $\frac{1}{n}X_{0,n}$ still converges a.e. to a T -invariant limit $\bar{X}$ taking values in $[-\infty, \infty)$.

Subadditive processes first arose in the work of Hammersley and Welsh (1965) on percolation theory. Here is an example. Let G be the graph with vertex set $\mathbb{Z}^2$ and with edges joining every pair of nearest neighbors. Let E denote the edge set and let $\{T_e : e \in E\}$ be non-negative integrable i.i.d. random variables. To each finite path P in G associate the “travel time” $T(P) = \sum_{e \in P} T_e$. For integers $m > n \geq 0$ let $X_{n,m}$ be the infimum of $T(P)$ over all paths P joining $(0, n)$ to $(0, m)$. This is a subadditive process with $0 \leq \gamma < \int T_e d\mu$ and it is not hard to see that the underlying endomorphism is ergodic. Thus Kingman's theorem yields the result that $\frac{1}{n}X_{0,n} \rightarrow \gamma$ a.e.

Suppose now that T is an ergodic automorphism of a probability space $(X, \mathcal{B}, \mu)$ and P is a function on X taking values in the space of $d \times d$ real matrices. Define $P_{n,m} = P(T^{m-1}x) \dots P(T^n x)$ and let $P_{n,m}(i, j)$ denote the i, j entry of $P_{n,m}$. Then $X_{n,m} = \log(\|P_{n,m}\|)$ (use any matrix norm) is a subadditive process so one obtains the first part of the following result of Furstenberg and Kesten (1960) originally proved by more elaborate methods. The second part can also be deduced from the subadditive theorem with a little more work. See Kingman (1976) for details and for some other applications of subadditive processes.

Theorem 25 (a) *Suppose $\int \log^+(|P|) d\mu < \infty$. Then $\|P_{0,n}\|^{1/n}$ converges a.e. to a finite limit.*

(b) *Suppose that for each i, j $P(i, j)$ is a strictly positive function such that $\log P(i, j)$ is integrable. Then the limit $p = \lim (P_{0,n}(i, j))^{1/n}$ exists a.e. and is independent of i and j .*

Partial results generalizing Kingman's theorem to the multiparameter case were obtained by Smythe (1976) and Nguyen (1979). In 1981 Akcoglu and Krengel (1981) obtained a definitive multi-parameter subadditive theorem. They consider an action $\{T_m\}$ of the semigroup $G = \mathbb{Z}_{\geq 0}^d$ by endomorphisms of a measure space $(X, \mathcal{B}, \mu)$. Using the standard total ordering $<$ of G an *interval* in G is any set of the form $\{k \in G : m < k < n\}$ for any $m < n \in G$. Let $\mathcal{I}$ denote the set of non-empty intervals. Reversing the direction of the inequality, they define a *superadditive* process as a collection of integrable functions $F_I, I \in \mathcal{I}$, such that

- (a) $F_I \circ T_m = F_{I+m}$,
- (b) $F_I \geq F_{I_1} + \dots + F_{I_k}$ whenever I is the disjoint union of $I_1, \dots, I_k$ and
- (c) $\gamma = \sup_{I \in \mathcal{I}} |I|^{-1} \int F_I d\mu < \infty$.

A sequence $\{I_n\}$ of sets in $\mathcal{I}$ is called *regular* if there is an increasing sequence I'_n such that $I_n \subset I'_n$ and $|I'_n| \leq C |I_n|$ for some constant C .

Theorem 26 (Akcoglu–Krengel) *Suppose F_I is a superadditive process and $\{I_n\}$ is regular. Then $\frac{1}{|I_n|} F_{I_n}$ converges a.e.*

$\{F_I\}$ is additive if the inequality in (b) is replaced by equality. In this case $F_I = \sum_{n \in I} f \circ T_n$ where f is an integrable function. Thus in the additive case the Akcoglu–Krengel result is a theorem about ordinary multi-dimensional ergodic averages, which is in fact a special case of an earlier result of Tempel'man (1972a) (see section “Amenable Groups” below).

Kingman's proof of Theorem 24 hinged on the existence of a certain (typically non-unique) decomposition for subadditive processes. Akcoglu and Krengel's proof of the multi-parameter result does not depend on a Kingman-

type decomposition, in fact they show that there is no such decomposition in general. They prove a weak maximal inequality

$$\mu\left\{\sup|I_n|^{-1}F_{I_n} > \lambda\right\} < \frac{C}{\lambda}\gamma, \quad (26)$$

where C is a constant depending only on the dimension, and show that this is sufficient to prove their result. In the case $d = 1$ the Akcoglu–Krengel argument provides a new and more natural proof of Kingman’s theorem, similar in spirit to Wiener’s arguments.

Akcoglu and Sucheston (1978) have proved a ratio ergodic theorem for subadditive processes with respect to a positive L_1 contraction, generalizing both the Chacón–Ornstein theorem and Kingman’s theorem.

In 1968 Oseledec (1968) proved his celebrated multiplicative ergodic theorem, which gives very precise information about the random matrix products studied by Furstenberg and Kesten. His theorem is an important tool for the study of Lyapunov exponents in differentiable dynamics, see notably Pesin (1977). If A is a $d \times d$ matrix let $\|A\| = \sup\{\|Ax\| : \|x\| = 1\}$ where $\|x\|$ is the Euclidean norm on $\mathbb{R}^n$.

Theorem 27 *Suppose T is an endomorphism of the probability space $(X, \mathcal{B}, \mu)$. Suppose P is a measurable function on X whose values are $d \times d$ real matrices such that $\int \log^+ \|P\| d\mu < \infty$ and let $P_n(x) = P(T^{n-1}x)P(T^{n-2}x) \dots P(x)$. Then there is a T -invariant subset X' of X with measure 1 such that for $x \in X'$ the following hold.*

- (a) $\lim_{n \rightarrow \infty} (P_n^*(x)P_n(x))^{\frac{1}{n}} = A(x)$ exists.
- (b) Let $0 \leq \exp \lambda_1(x) \leq \exp \lambda_2(x) \leq \dots \leq \exp \lambda_r(x)$ be the distinct eigenvalues of $A(x)$ ($r = r(x)$ may depend on x and λ_1 may be $-\infty$) with multiplicities $m_1(x), \dots, m_r(x)$. Let $E_i(x)$ be the eigenspace corresponding to $\exp(\lambda_i(x))$ and set.

$$F_i(x) = E_1(x) + \dots + E_i(x).$$

Then for each $u \in F_i(x) \setminus F_{i-1}(x)$

$$\lim \frac{1}{n} \log \|P_n(x)u\| = \lambda_i(x).$$

- (c) The functions m_i and λ_i are T -invariant.
- (d) If T is ergodic, $\det P(x) = 1$ a.e. and

$$\limsup_n \frac{1}{n} \int \log \|P_n\| d\mu > 0$$

then the λ_i are constants, $\lambda_1 < 0$ and $\lambda_r > 0$.

Raghuathan (1979) gave a much shorter proof of Oseledec’s theorem, valid for matrices with entries in a locally compact normed field. He showed that it could be reduced to the Furstenberg–Kesten theorem by considering the exterior powers of P . Ruelle (1982) extended Oseledec’s theorem to the case where P takes values in the set of bounded operators on a Hilbert space. Walters (1993) has given a proof (under slightly stronger hypotheses) which avoids the matrix calculations and tools from multilinear algebra used in other proofs.

Entropy and the Shannon–McMillan–Breiman Theorem

The notion of entropy was introduced by Shannon in his landmark work (Shannon 1948) which laid the foundations for a mathematical theory of information. Suppose $(X, \mathcal{B}, \mu)$ is a probability space, P is a finite measurable partition of X and T is an automorphism of $(X, \mathcal{B}, \mu)$. $P(x)$ denotes the atom of P containing x . The entropy of P is

$$\begin{aligned} h(P) &= - \sum_{p \in P} \mu(p) \log(\mu(p)) \\ &= - \int \log(\mu(P(x))) d\mu(x) \geq 0. \end{aligned} \quad (27)$$

$-\log(\mu(A))$ may be viewed as a quantitative measure of the amount of information contained in the statement that a randomly chosen $x \in X$

happens to belong to A . So $h(P)$ is the expected information if one is about to observe which atom of P a randomly chosen point falls in. See Billingsley (1965) for more motivation of this concept. See also the article in this collection on entropy by J. King or any introductory book on ergodic theory, e.g. Petersen (1989).

If P and Q are partitions $P \vee Q$ denotes the common refinement which consists of all sets $p \cap q$, $p \in P$, $q \in Q$. It is intuitive and not hard to show that $h(P \vee Q) \leq h(P) + h(Q)$. Now let $P_0^n = \bigvee_{i=0}^{n-1} T^{-i} P$ and $h_n = h(P_0^n)$. The subadditivity of entropy implies that $h_{n+m} \leq h_n + h_m$, so by a well-known elementary lemma the limit

$$h(P, T) = \lim_n \frac{h_n}{n} = \inf_n \frac{h_n}{n} \geq 0 \quad (28)$$

exists.

If one thinks of $P(x)$ as a measurement performed on the space X and Tx as the state succeeding x after one second has elapsed then $h(P, T)$ is the expected information per second obtained by repeating the experiment every second for a very long time. See (Rudolph 1990) for an alternative and very useful approach to $h(P, T)$ via *name-counting*.

The following result, which is known as the Shannon–McMillan–Breiman theorem, has proved to be of fundamental importance in ergodic theory, notably, for example, in the proof of Ornstein’s celebrated isomorphism theorem for Bernoulli shifts (Ornstein 1970).

Theorem 28 *If T is ergodic then*

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log(\mu(P_0^{n-1}(x))) = h(P, T) \quad (29)$$

a.e. and in L_1 -norm.

In other words, the actual information obtained per second by observing x over time converges to the constant $h(P, T)$, namely the limiting expected information per second. Shannon (1948) formulated Theorem 28 and proved convergence in probability. McMillan (1953) proved L_1

convergence and Breiman (1957) obtained the a.e. convergence.

The original proofs of a.e. convergence used the martingale convergence theorem, were not very intuitive and did not generalize to $\mathbb{Z}^n$ -actions, where the martingale theorem is not available. Ornstein and Weiss (1983) found a beautiful and more natural argument which bypasses the martingale theorem and allows generalization to a class of groups which includes $\mathbb{Z}^n$.

Amenable Groups

Let G be any countable group and $T = \{T_g\}$ an action of G by automorphisms of a probability space. Suppose σ is a complex measure on G , that is, $\{\sigma(g)\}_{g \in G}$ is an absolutely summable sequence. Let T_g act on functions via $T_g f = f \circ T_g$, so T_g is an isometry of L_p for every $1 \leq p \leq \infty$. Let $\sigma(T) = \sum_{g \in G} \sigma(g) T_g$. A very general framework for formulating ergodic theorems is to consider a sequence $\{\sigma_n\}$ and ask whether $\sigma_n(T)f$ converges, a.e. or in mean, for f in some L_p space. When $\sigma_n(T)f$ converges for all actions T and all f in L_p in p -norm or a.e. then one says that σ_n is *mean* or *pointwise good* in L_p . When the σ_n are probability measures it is natural to call such results weighted ergodic theorems and this terminology is retained for complex σ as well.

Birkhoff’s theorem says that if $G = \mathbb{Z}$ and σ_n is the normalized counting measure on $\{0, 1, \dots, n-1\}$ then $\{\sigma_n\}$ is pointwise good in L_1 . This section will be concerned only with sequences $\{\sigma_n\}$ such that σ_n is normalized counting measure on a finite subset $F_n \subset G$ so one speaks of mean or pointwise good sequences $\{F_n\}$. A natural condition to require of $\{F_n\}$, which will ensure that the limiting function is invariant, is that it be asymptotically (left) invariant in the sense that

$$\frac{|g F_n \Delta F_n|}{|F_n|} \rightarrow 0 \quad \forall g \in G. \quad (30)$$

Such a sequence is called a *Følner sequence* and a group G is *amenable* if it has a Følner sequence. As in most of this article G is restricted to be a discrete countable group for simplicity but

most of the results to be seen actually hold for a general locally compact group.

Amenability of G is equivalent to the existence of a finitely additive left invariant probability measure on G . It is not hard to see that any Abelian, and more generally any solvable, group is amenable. On the other hand the free group F_2 on two generators is not amenable. See Paterson (1988) for more information on amenable groups. The Følner property by itself is enough to give a mean ergodic theorem.

Theorem 29 *Any Følner sequence is mean good in L_p for $1 \leq p < \infty$.*

The proof of this result is rather similar to the proof of Theorem 3. In fact Theorem 29 is only a special case of quite general results concerning amenable semi-groups acting on abstract Banachspaces. See the book of Paterson (1988) for more on this.

Turning to pointwise theorems, the Følner condition alone does not yield a pointwise theorem, even when $G = \mathbb{Z}$ and the F_n are intervals. For example Akcoglu and del Junco (1975) have shown that when $G = \mathbb{Z}$ and $F_n = [n, n + \sqrt{n}] \cap \mathbb{Z}$ the pointwise ergodic theorem fails for any aperiodic T and for some characteristic function f . See also del Junco and Rosenblatt (1979).

The following pointwise result of Tempelman (1972a) is often quoted. A Følner sequence $\{F_n\}$ is called *regular* if there is a constant C such that $|F_n^{-1}F_n| \leq C |F_n|$ and there is an increasing sequence F'_n such that $F_n \subset F'_n$ and $|F'_n| \leq C |F_n|$.

Theorem 30 *Any regular Følner sequence is pointwise good in L_1 .*

In case the F_n are intervals in $\mathbb{Z}^n$ this result can be proved by a variant of Wiener's covering argument and in the general case by an abstraction thereof. The condition $|F_n^{-1}F_n| \leq C |F_n|$ captures the property of rectangles which is needed for the covering argument. Emerson (1974) independently proved a very similar result.

The work on ergodic theorems for abstract locally compact groups was pioneered by Calderón (1953) who built on Wiener's methods. The main result in this paper is somewhat technical but it already contains the germ of Tempelman's theorem. Other ergodic theorems for amenable groups, whose main interest lies in the case of continuous groups, include Tempelman (1967), Renaud (1971), Greenleaf (1973) and Greenleaf and Emerson (1974). The discrete versions of these results are all rather close to Tempelman's theorem.

Among pointwise theorems for discrete groups Tempelman's result was essentially the best available for a long time. It was not known whether every amenable group had a Følner sequence which is pointwise good for some L_p . In 1988 Shulman (1988) introduced the notion of a *tempered* Følner sequence $\{F_n\}$, namely one for which

$$\left| \bigcup_{i < n} F_i^{-1} F_n \right| < C |F_n|, \quad (31)$$

for some constant C . The advantage of the tempered condition is that any Følner sequence has a tempered subsequence, and in particular any amenable group has a tempered Følner sequence.

Shulman proved a maximal inequality in L_2 for such F_n which implies that $\{F_n\}$ is pointwise good in L_2 . An account of this work may be found in Section 5.6 of Tempelman's book (1992).

Lindenstrauss (1999) was able to extend the result to L_1 .

Theorem 31 *Any tempered Følner sequence is pointwise good in L_1 .*

The key new idea in his proof is to use a probabilistic argument to establish a covering lemma. In the discrete case Ornstein and Weiss (2003) have given a non-probabilistic proof of Lindenstrauss's covering lemma. Lindenstrauss also generalizes the a.e. convergence in the Shannon–McMillan–Breiman theorem to this setting. L_1 convergence was already established by Kieffer (1975) in 1975.

Subsequence and Weighted Theorems

In this section G and T are as in the previous section and σ_n is a sequence of complex measures on G .

This section will be concerned with conditions on σ_n that ensure that it is mean or pointwise good. For the most part G will be $\mathbb{Z}$.

Hopf's ergodic theorem gives a class of examples for free. Choose any probability measure λ on G , define the operator $T_\lambda = \lambda(T)$ and observe that $(T_\lambda)^n = T_{\lambda^{*n}}$, where λ^{*n} denotes the convolution power. Since T_λ satisfies the hypotheses of Hopf's theorem it follows that the sequence $\sigma_n = \frac{1}{n} \sum_{i=0}^{n-1} \lambda^{*i}$ is pointwise good in L_1 .

Another sort of example is given by choosing a sequence g_n in G and letting $\sigma_n = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{g_n}$. If one has convergence for such a sequence one speaks of a *subsequence ergodic theorem*. This section will mainly focus on the case $G = \mathbb{Z}$ and sequences which are the increasing enumeration of a subset $S \subset \mathbb{N}$. Given $S \subset \mathbb{N}$ write $\sigma_{S,n}$ for the corresponding probabilities σ_n and say S is good if $\sigma_{S,n}$ is good.

Perhaps the first subsequence ergodic theorem is due to Blum and Hanson (1960) who proved that an automorphism T of a probability space is *strongly mixing* if and only if $\frac{1}{n} \sum_{i=0}^{n-1} T^{m_i} f$ converges in L_2 norm for every $f \in L_2$ and every increasing $\{m_i\}$. Strong mixing means that $\mu(T^{-n}A \cap B) \rightarrow \mu(A)\mu(B)$. In 1969 Brunel and Keane (1969) proved the first pointwise subsequence ergodic theorem.

Theorem 32 *Suppose that T is a translation on a compact Abelian group G with Haar measure λ , $g \in G$, and $E \subset G$ is any Borel set with $\lambda(E) > 0$ and $\lambda(\partial E) = 0$. Let $S = \{i > 0 : T^i g \in E\}$. Then S is pointwise good in L_1 .*

Krengel (1971) constructed the first example of a sequence $S \subset \mathbb{N}$ which is pointwise universally bad, in the strong sense that for any aperiodic T the a.e. convergence of $\sigma_{S,n}(T)f$ fails for some characteristic function f . Bellow (1983) proved that any lacunary sequence (meaning $a_{n+1} > ca_n$ for some $c > 1$) is pointwise universally bad in L_1 .

Later Akcoglu et al. (1996), were able to show that for lacunary sequences $\{\sigma_{S,n}\}$ is even *strongly sweeping out*. A sequence $\{\sigma_n\}$ of probability measures on $\mathbb{Z}$ is said to be strongly sweeping out if for any ergodic T and for all $\delta > 0$ there is a characteristic function f with $\int f d\mu < \delta$ such that $\limsup \sigma_n(T)f = 1$ a.e. It is not difficult to show that if $\{\sigma_n\}$ is strongly sweeping out then there are characteristic functions f such that $\liminf \sigma_n(T)f = 0$ and $\limsup \sigma_n(T)f = 1$. Thus for lacunary sequences the ergodic theorem fails in the worst possible way.

Bellow and Losert (1985) gave the first example of a sequence $S \subset \mathbb{Z}$ of density 0 which is universally good for pointwise convergence, answering a question posed by Furstenberg. They construct an S which is pointwise good in L_1 . This paper also contains a good overview of the progress on weighted and subsequence ergodic theorems at that time.

Weyl's theorem on uniform distribution (Theorem 9) suggests the possibility of an ergodic theorem for the sequence $\{n^2\}$. It is not hard to see that $\{n^2\}$ is mean good in L_2 . In fact the spectral theorem and the dominated convergence theorem show that it is enough to prove that the L_∞ -bounded sequence of functions $\frac{1}{n} \sum_{i=0}^{n-1} z^{n^2}$ on the unit circle converges at each point z of the unit circle. When z is not a root of unity the sequence converges to 0 by Weyl's result and when z is a root of unity the convergence is trivial because $\{z^{n^2}\}$ is periodic. In 1987 Bourgain (1987, 1988d) proved his celebrated pointwise ergodic theorem for polynomial subsequences.

Theorem 33 *If p is any polynomial with rational coefficients taking integer values on the integers then $S = \{p(n)\}$ is pointwise good in L_2 .*

The first step in Bourgain's argument is to reduce the problem of proving a maximal inequality to the case of the shift map on the integers, via Calderón's transfer principle. Then the problem is transferred to the circle by using Fourier transforms. At this point the problem becomes a very delicate question about exponential sums and a whole arsenal of tools is brought to bear. See

Rosenblatt and Wierdl (1995) and Quas and Wierdl (Bergelson 2006) (Appendix B) for nice expositions of Bourgain's methods.

Bourgain subsequently improved this to all L_p , $p > 1$ and also extended it to sequences $\{[q(n)]\}$ where now q is an arbitrary real polynomial and $[\cdot]$ denotes the greatest integer function. He also announced that his methods can be used to show that the sequence of primes is pointwise good in L_p for any $p > \frac{1+\sqrt{3}}{2}$. Wierdl (1988) soon extended the result for primes to all $p > 1$.

Theorem 34 The primes are pointwise good in L_p for $p > 1$.

It has remained a major open question for quite some time whether any of these results hold for $p = 1$. In 2005 there appeared a preprint of Mauldin and Buczolicz (2005), which remains unpublished, showing that polynomial sequences are L_1 -universally bad.

Another major result of Bourgain's is the so-called return times theorem (Bourgain 1988e). A simplification of Bourgain's original proof was published jointly with Furstenberg, Katznelson and Ornstein as an appendix to an article (Bourgain 1989b) of Bourgain. To state it let us agree to say that a sequence of complex numbers $\{a(n)\}_{n \geq 0}$ has property P if the sequence of complex measures $\sigma_n = \frac{1}{n} \sum_{i=0}^{n-1} a(i) \delta_i$ has property P , where δ_i denotes the point mass at i .

Theorem 35 (Bourgain) Suppose T is an automorphism of a probability space $(X, \mathcal{B}, \mu)$, $1 \leq p$, $q \leq \infty$ are conjugate exponents and $f \in L_p(\mu)$. Then for almost all x the sequence $\{f(T^n x)\}$ is pointwise good in L_q .

Applying this to characteristic functions $f = 1_E$ one sees that the return time sequence $\{i > 0 : S^i x \in E\}$ is good for pointwise convergence in L_1 . Theorem 32 is a very special case. It is also easy to see that Theorem 35 contains the Wiener–Wintner theorem.

In 1998 Rudolph (1998) proved a far-reaching generalization of the return times theorem using the technique of *joinings*. For an introduction to joinings see the article by de la Rue in this collection

and also Thouvenot (1995), Glasner (2003) and Rudolph's book (1990). Rudolph's result concerns the convergence of multiple averages

$$\frac{1}{N} \sum_{n=0}^{N-1} \prod_{j=1}^k f_j(T_j^n x) \quad (32)$$

where each T_j is an automorphism of a probability space $(X_j, \mathcal{B}_j, \mu_j)$ and the f_j are L_∞ functions. The point is that the convergence occurs whenever each $x_j \in X'_j$, sets of measure one which may be chosen sequentially for $j = 1, \dots, k$ without knowing what T_i or f_i are for any $i > j$. He actually proves something stronger, namely he identifies an intrinsic property of a sequence $\{a_i\}$, which he calls *fully generic*, such that the following hold.

- (a) The constant sequence $\{1\}$ is fully generic.
- (b) If $\{a_i\}$ is fully generic then for any T and $f \in L_\infty$ the sequence $a_i f(T^i x)$ is fully generic for almost all x .
- (c) Fully generic implies pointwise good in L_∞ .

The definition of fully generic will not be quoted here as it is somewhat technical.

For a proof of the basic return times theorem using joinings see Rudolph (1994). Assani, Lesigne and Rudolph (1995) took a first step towards the multiple theorem, a Wiener–Wintner version of the return times theorem. Also Assani (2000) independently gave a proof of Rudolph's result in the case when all the T_j are weakly mixing.

Ornstein and Weiss (1992) have proved the following version of the return times theorem for abstract discrete groups. As with $\mathbb{Z}$, let us say that a sequence $\{a_g\}_{g \in G}$ of complex numbers has property P for $\{F_n\}$ if the sequence $\sigma_n = \frac{1}{|F_n|} \sum_{g \in F_n} a(g) \delta_g$ of complex measures has property P .

Theorem 36 Suppose that the increasing Følner sequence $\{F_n\}$ satisfies the Tempelman condition $\sup_n |F_n^{-1} F_n| / |F_n| < \infty$ and $\cup F_n = G$. If $b \in L_\infty$ then for a.a. x the sequence $\{b(T_g x)\}$ is pointwise good in L_1 for $\{F_n\}$.

Recently Demeter, Lacey, Tao and Thiele (2008) have proved that the return times theorem

remains valid for any $1 < p \leq \infty$ and $q \geq 2$. On the other hand Assani, Buczolic and Mauldin (2005) showed that it fails for $p = q = 1$.

Bellow, Jones and Rosenblatt have a series of papers (Bellow et al. 1989, 1990, 1992, 1994) studying general weighted averages associated to a sequence σ_n of probability measures on $\mathbb{Z}$, and, in some cases, more general groups. The following are a few of their results (Bellow et al. 1990). is concerned with $\mathbb{Z}$ -actions and *moving block* averages given by $\sigma_n = m_{I_n}$, where the I_n are finite intervals and m_I denotes normalized counting measure on I . They resolve the problem completely, obtaining a checkable necessary and sufficient condition for such a sequence to be pointwise good in L_1 .

Bellow et al. (1992) gives sufficient conditions on a sequence σ_n for it to be pointwise good in L_p , $p > 1$, via properties of the Fourier transforms $\hat{\sigma}_n$. A particular consequence is that if $\lim_{n \rightarrow \infty} \sum_{k \in \mathbb{Z}} |\sigma_n(k) - \sigma_n(k-1)| = 0$ then $\{\sigma_n\}$ has a subsequence which is pointwise good in L_p , $p > 1$. In (Bellow et al. 1994) they obtain convergence results for sequences $\sigma_n = \sigma^n$, the convolution powers of a probability measure σ . A consequence of one of their main results is that if the expectation $\sum_{k \in \mathbb{Z}} k \sigma(k)$ is zero, the second moment $\sum_{k \in \mathbb{Z}} k^2 \sigma(k)$ is finite and σ is *aperiodic* (its support is not contained in any proper coset in $\mathbb{Z}$) then σ^n is pointwise good in L_p for $p > 1$.

Bellow and Calderón (1999) later showed that this last result is valid also for $p = 1$. This is a consequence of the following sufficient condition For a sequence T to satisfy a weak L_1 inequality. Given an automorphism of a probability space $(X, \mathcal{B}, \mu)$ let $Mf = \sup |\sigma_n(T)f(x)|$ be the maximal operator associated to $\{\sigma_n\}$.

Theorem 37 (Bellow and Calderón) *Suppose there is an $\alpha \in (0, 1]$ and $C > 0$ such that for each $n > 1$ one has*

$$|\sigma_n(x+y) - \sigma_n(x)| \leq C \frac{|y|^\alpha}{|x|^{1+\alpha}} \text{ for all } x, y \in \mathbb{Z}$$

such that $0 < 2|y| \leq |x|$

Then there is a constant D such that

$$\mu\{Mf > \lambda\} \leq \frac{D}{\lambda} \|f\|_1 \text{ for all } T, f \in L_1(\mu)$$

and $\lambda > 0$.

Ergodic Theorems and Multiple Recurrence

Suppose $S \subset \mathbb{N}$. The *upper density* of S is

$$\bar{d}(S) = \limsup_n \frac{|S \cap [1, n]|}{n}. \quad (33)$$

and the *density* $d(S)$ is the limit of the same quantity, if it exists. In 1975 Szemerédi (1975) proved the following celebrated theorem, answering an old question of Erdős and Turán.

Theorem 38 *Any subset of $\mathbb{N}$ with positive upper density contains an arithmetic progression of length k for each $k \geq 1$.*

This result has a distinctly ergodic-theoretic flavor. Letting T denote the shift map on $\mathbb{Z}$, it says that for each k there is an n such that $S' = \cap_{i=1}^k T^{-in}S$ is non-empty. In fact the result gives more: there is an n for which $\bar{d}(S') > 0$. In this light Szemerédi's theorem becomes a multiple recurrence theorem for the shift map on $\mathbb{N}$, equipped with the invariant “measure-like” quantity $\bar{d}$. Of course $\bar{d}$ is not even finitely additive so it is *not* a measure. d , however, is at least finitely additive, when defined, and $d(\mathbb{N}) = 1$.

This point of view suggests the following multiple recurrence theorem.

Theorem 39 *Suppose T is an automorphism of a probability space $(X, \mathcal{B}, \mu)$, $\mu(B) > 0$ and $k \geq 1$. Then there is an $n > 0$ such that $\mu(\cap_{i=1}^k T^{in}B) > 0$.*

In 1977, Furstenberg (1977) proved the following ergodic theorem which implies the multiple recurrence theorem. He also established a general *correspondence principle* which puts the shaky analogy between the multiple recurrence theorem

and Szemerédi's theorem on a firm footing and allows each to be deduced from the other. Thus he obtained an ergodic theoretic proof of Szemerédi's combinatorial result.

Theorem 40 *Suppose T is an automorphism of a probability space $(X, \mathcal{B}, \mu)$, $f \in L_\infty$, $f \geq 0$, $\int f d\mu > 0$ and $k \geq 1$. Then*

$$\liminf_N \frac{1}{N} \sum_{n=0}^{N-1} \int \prod_{i=1}^k T^{in} f d\mu > 0. \quad (34)$$

Furstenberg's result opened the door to the study of so-called ergodic Ramsey theory which has yielded a vast array of deep results in combinatorics, many of which have no non-ergodic proof as yet. The focus of this article is not on this direction but the reader is referred to Furstenberg's book (1981) for an excellent introduction and to Bergelson (1996, 2006) for surveys of later developments. There is also the article by Frantzikinakis and McCutcheon in this collection.

Furstenberg's proof relies on a deep structure theorem for a general automorphism which was also developed independently by Zimmer (1976a, b) in a more general context. A *factor* of T is any sub- σ -algebra $\mathcal{F} \subset \mathcal{B}$ such that $T(\mathcal{F}) = \mathcal{F}$. (It is more accurate to think of the factor as the action of T on the measure space $(X, \mathcal{F}, \mu|_{\mathcal{F}})$). The structure theorem asserts that there is a transfinite increasing sequence of factors $\{\mathcal{F}_\alpha\}$ of T such that the following conditions hold.

- (a) $\mathcal{F}_{\alpha+1}$ is compact relative to $\mathcal{F}_\alpha$.
- (b) $\mathcal{F}_\alpha = \bigvee_{\beta < \alpha} \mathcal{F}_\beta$ whenever α is a limit ordinal.
- (c) $\mathcal{B}$ is weakly mixing relative to $\bigvee \mathcal{F}_\alpha$.

($\bigvee$ denotes the σ -algebra generated by a collection of σ -algebras.) $\bigvee_\alpha \mathcal{F}_\alpha$ is called the *maximal distal factor* of T and if $\bigvee \mathcal{F}_\alpha = \mathcal{B}$ then T is called *distal*. The definitions of the relative properties in (a) and (c) above are somewhat technical

so only their absolute versions (i. e. relative to the trivial σ -algebra $\{\emptyset, X\}$) will be described here.

T is said to be compact if for each $f \in L_2$ the orbit $\{T^i f : i \in \mathbb{Z}\}$ is pre-compact in the norm topology of L_2 . This turns out to be equivalent to the statement that T is a translation on a compact Abelian group endowed with its Haar measure. The property of weak mixing is a fundamental notion in ergodic theory which has many equivalent definitions. The most appropriate for our purposes is that T is weakly mixing if it has no compact factors. This turns out to be equivalent to the ergodicity of $T \times T$ acting on the product measure space $(X, \mathcal{B}, \mu) \times (X, \mathcal{B}, \mu)$.

The verification of 37 in the case of compact T is rather easy. In this case it is not hard to prove that for any $f \in L_\infty$ and $\epsilon > 0$ the set $\{n \in \mathbb{Z} : \|T^n f - f\|_2 < \epsilon\}$ has bounded gaps and (34) follows easily. In the case when T is weakly mixing (34) is a consequence of the following theorem which Furstenberg proves in (1977) (as a warm-up for its much harder relative version).

Theorem 41 *If T is weakly mixing and $f_1, f_2, \dots, f_k$ are L_∞ functions then*

$$\lim_N \frac{1}{N} \sum_{n=0}^{N-1} \int \prod_{i=1}^k T^{in} f_i d\mu = \prod_{i=1}^k \int f_i \quad (35)$$

Later Bergelson (1987) showed that the result can be obtained easily by an induction argument using the following Hilbert space generalization of van der Corput's lemma.

Theorem 42 (Bergelson) *Suppose $\{x_n\}$ is a bounded sequence of vectors in Hilbert space such that for each $h > 0$ one has $\frac{1}{N} \sum_{n=0}^{N-1} \langle x_{n+h}, x_n \rangle \rightarrow 0$ as $N \rightarrow \infty$. Then $\left\| \frac{1}{N} \sum_{n=0}^{N-1} x_n \right\| \rightarrow 0$.*

Ryzhikov has also given a beautiful short proof of Theorem 41 using joinings (Ryzhikov 1994). Bergelson's van der Corput lemma and variants of

it have been a key tool in subsequent developments in ergodic Ramsey theory and in the convergence results to be discussed in this section. Bergelson (1987) used it to prove the following mean ergodic theorem for weakly mixing automorphisms.

Theorem 43 *Suppose T is weakly mixing, $f_1, \dots, f_k$ are L_∞ functions and $p_1, \dots, p_k$ are polynomials with rational coefficients taking integer values on the integers such that no $p_i - p_j$ is constant for $i \neq j$. Then*

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=1}^k T^{p_i(n)} f_i - \prod_{i=1}^k \int f_i d\mu \right\| = 0. \quad (36)$$

Theorems 40 and 41 immediately raise the question of convergence of the multiple averages $\frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=1}^k T^{n p_i} f_i$ for a general T . Several authors obtained partial results on the question of mean convergence. It was finally resolved only recently by Host and Kra (2005b), who proved the following landmark theorem.

Theorem 44 *Suppose $f_1, f_2, \dots, f_k \in L_\infty$. Then there is a $g \in L_\infty$ such that*

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=1}^k T^{n p_i} f_i - g \right\|_2 = 0. \quad (37)$$

Independently and somewhat later Ziegler (2007) obtained the same result by somewhat different methods. Furstenberg had already established Theorem 44 for $k = 2$ in (Furstenberg 1977). It was proved for $k = 3$ in the case of a totally ergodic T by Conze and Lesigne (1988) and in general by Host and Kra (2001). It can also be obtained using the methods developed by Furstenberg and Weiss (1996). In this paper Furstenberg and Weiss proved a result for polynomial powers of T in the case $k = 2$. They also formalized the key notion of a *characteristic factor*. A factor C of T is said to be characteristic for the averages (37) if, roughly speaking, the L_2

limiting behavior of the averages is unchanged when any one of the f_i 's is replaced by its conditional expectation on C . This means that the question of convergence of these averages may be reduced to the case when f_i are all C -measurable. So the problem is to find the right (smallest) characteristic factor and prove convergence for that factor.

The importance of characteristic factors was already apparent in Furstenberg's original paper (1977), where he showed that the maximal distal factor is characteristic for the averages (37). In fact he showed that for a given k a k -step distal factor is characteristic. (An automorphism is k -step distal if it is the top rung in a k -step ladder of factors as in the Furstenberg–Zimmer structure-theorem.) It turns out, though, that the right characteristic factor for (37) is considerably smaller. In their seminal paper (Conze and Lesigne 1988) Conze and Lesigne identified the characteristic factor for $k = 3$, now called the Conze–Lesigne factor. As shown in (Host and Kra 2005b; Ziegler 2007), the characteristic factor for a general k is (isomorphic to) an inverse limit of k -step nilflows. A k -step nilflow is a compact homogeneous space N/Γ of a k -step nilpotent Lie group N , endowed with its unique left-invariant probability measure, on which T acts via left translation by an element of N . Ergodic properties of nilflows have been studied for some time in ergodic theory, for example in Parry (1969). In this way the problem of L_2 -convergence of (37) is reduced to the case when T is a nilflow. In this case one has more: the averages converge pointwise by a result of Leibman (2005a) (See also Ziegler (2005)).

There have already been a good many generalizations of (37). Host and Kra (2005a), Frantzikinakis and Kra (2005b, 2006), and Leibman (2005a) have proved results which replace linear powers of T by polynomial powers. In increasing degrees of generality Conze and Lesigne (1988), Frantzikinakis and Kra (2005a) and Tao (2008) have obtained results which replace the maps $T, T^2, \dots, T^k$ in (37) by commuting maps $T_1, \dots, T_k$. Bergelson and Leibman (2002, 2004) have obtained results, both positive and negative, in the case of two non-commuting maps.

In the direction of pointwise convergence the only general result is the following theorem of Bourgain (1990) which asserts pointwise convergence in the case $k = 2$.

Theorem 45 *Suppose S and T are powers of a single automorphism R and $f, g \in L_\infty$. Then $\frac{1}{N} \sum_{n=1}^N f(T^n x)g(S^n x)$ converges a.e.*

When T is a K-automorphism Derrien and Lesigne (1996) have proved that the averages (35) converge pointwise to the product of the integrals, even with polynomial powers of T replacing the linear powers.

Gowers (2001) has given a new proof of Szemerédi's theorem by purely finite methods using only harmonic analysis on $\mathbb{Z}_n$. His results give better quantitative estimates in the statement of finite versions of Szemerédi's theorem. Although his proof contains no ergodic theory it is to some extent guided by Furstenberg's approach.

This section would be incomplete without mentioning the spectacular recent result of Green and Tao (2004) on primes in arithmetic progression and the subsequent extensions of the Green-Tao theorem due to Tao (2005) and Tao and Ziegler (2006).

Rates of Convergence

There are many results which say in various ways that, in general, there are no estimates for the rate of convergence of the averages $A_n f$ in Birkhoff's theorem. For example there is the following result of Krengel (1978).

Theorem 46 *Suppose $\lim_{n \rightarrow \infty} c_n = \infty$ and T is any ergodic automorphism of a probability space. Then there is a bounded measurable f with $\int f d\mu = 0$ such that $\limsup c_n A_n f = \infty$ a.e.*

See Part 1 of Derriennic (2006) for a selection of other results in this direction. In spite of these negative results one can obtain quantitative estimates by reformulating the ergodic

theorem in various ways. Bishop (1967) proved the following result which is purely finite and constructive in nature and evidently implies the a.e. convergence in Birkhoff's theorem. If $y = (y_1, \dots, y_n)$ is a finite sequence of real numbers and $a < b$, an *upcrossing* of y over $[a, b]$ is a minimal integer interval $[k, l] \subset [1, n]$ satisfying $y_k < a$ and $y_l > b$.

Theorem 47 *Suppose T is an automorphism of the probability space $(X, \mathcal{B}, \mu)$. Let $U(n, a, b, f, x)$ be the number of upcrossings of the sequence $A_0 f(x), \dots, A_n f(x)$ over $[a, b]$. Then for every n*

$$\mu\{x : U(n, a, b, f, x) > k\} \leq \frac{\|f\|_1}{(b-a)k}. \quad (38)$$

Ivanov (1996a) has obtained the following stronger upcrossing inequality for an arbitrary positive measurable f , which also implies Birkhoff's theorem.

Theorem 48 *For any positive measurable f and $0 < a < b$*

$$\mu\{x : U(n, a, b, f, x) > k\} \leq \left(\frac{a}{b}\right)^k. \quad (39)$$

Note the exponential decay and the remarkable fact that the estimate does not depend on f . Ivanov has also obtained the following result (Theorem 23 in (Kachurovskii 1996a)) about *fluctuations* of $A_n f$. An ϵ -*fluctuation* of a real sequence $y = (y_1, \dots, y_n)$ is a minimal integer interval $[k, l]$ satisfying $|y_k - y_l| \geq \epsilon$. If $f \in L_1(\mathbb{R})$ let $F(\epsilon, f, x)$ be the number of ϵ -fluctuations of the sequence $\{A_n f(x)\}_{n=0}^\infty$.

Theorem 49

$$\mu\{x : F(\epsilon, f, x) \geq k\} \leq C \frac{(\log k)^{\frac{1}{2}}}{k} \quad (40)$$

where C is a constant depending only on $\|f\|_1/\epsilon$. If $f \in L_\infty$ then

$$\mu\{x : F(\epsilon, f, x) \geq k\} \leq A e^{-Bk}, \quad (41)$$

where A and B are constants depending only on $\|f\|_\infty/\epsilon$.

See Kachurovskii's survey (1996a) of results on rates of convergence in the ergodic theorem and also the article of Jones et al. (1998) for more results on oscillation type inequalities.

Under special assumptions on f and T it is possible to give more precise results on speed of convergence. If the sequence $T^i f$ is independent then there is a vast literature in probability theory giving very precise results, for example the central limit theorem and the law of the iterated logarithm (see, for example, (Breiman 1968)). See the surveys of Derrien (2006) and of Merlevède, Peligrad and Utev (2006) for results on the central limit theorem for dynamical systems.

Ergodic Theorems for Non-amenable Groups

Guivarc'h (1969) was the first to prove an ergodic theorem for a general pair of non-commuting unitary operators. Much work has been done in the last 15 years on mean and pointwise theorems for non-amenable groups. See the extensive survey by Nevo (2006). Here is just one result of Nevo (1994) as a sample. Suppose G is a discrete group and $\{\sigma_n\}$ is a sequence of probability measures on G . Say that $\{\sigma_n\}$ is *mean ergodic* if for any unitary representation π of G on a Hilbert space $\mathcal{H}$ and any $x \in \mathcal{H}$ one has $\sum_{g \in G} \sigma_n(g) \pi(g)x$ converges in norm to the projection of x on the subspace of π -invariant vectors. Say that $\{\sigma_n\}$ is *pointwise ergodic* if for any measure preserving action T of G on a probability space and $f \in L_2$ one has $\sum_{g \in G} \sigma_n(g) T_g f$ converges a.e. to the projection of f on the subspace of T -invariant functions.

Theorem 50 *Let G be the free group on k generators, $k \geq 1$. Let τ_n be the normalized counting measure on the set of elements whose word length (in terms of the generators and their inverses) is n . Let $\sigma_n = (\tau_n + \tau_{n+1})/2$ and $\sigma'_n = \frac{1}{n} \sum_{i=1}^n \sigma_i$. Then σ_n and σ'_n are mean and pointwise ergodic but τ_n is not mean ergodic.*

Future Directions

This section will be devoted to a few open problems and some questions. Many of these have been suggested by my colleagues acknowledged in the introduction.

The topic of convergence of Furstenberg's multiple averages has seen some remarkable achievements in the last few years and will likely continue to be vital for some time to come. The question of pointwise convergence of multiple averages is completely open, beyond Bourgain's result (Theorem 45). Even extending Bourgain's result to three different powers of R or to any two commuting automorphisms S and T would be a very significant achievement. Another natural question is whether one has pointwise convergence of the averages of the sequence $f(T^n(x))g(T^{n^2}(x))$, which are mean convergent by the result of Furstenberg and Weiss (1996) (which is now subsumed in much more general results).

A long-standing open problem relating to Akcoglu's ergodic theorem (Theorem 23) for positive contractions of L_p is whether it can be extended to power-bounded operators T (this means that $\|T\|^n$ is bounded). It is also an open question whether it extends to non-positive contractions, excepting the case $p = 2$ where Burkholder (1962) has shown that it fails.

There are many natural questions in the area of subsequence and weighted ergodic theorems. For example, which of Bourgain's several pointwise theorems can be extended to L_1 ? Are there other natural subsequences of an arithmetic character which have density 0 and for which an ergodic theorem is valid, either mean or pointwise, and in what L_p spaces might such theorems be valid? Are there higher dimensional analogues?

Since lacunary sequences are bad, to have any sort of pointwise theorem $l_n = \log \frac{a_n}{a_{n+1}}$ must get close to 0 and for simplicity let us assume that $\lim l_n = 0$. How fast must the convergence be in order to get an ergodic theorem? Jones and Wierdl (1994) have shown that if $l_n > (\log n)^{-\frac{1}{2}+\epsilon}$ then

the pointwise ergodic theorem fails in L_2 while Jones, Lacey and Wierdl (1999) have shown that an only slightly faster rate permits a sequence which is pointwise good in L_2 . How well or badly does the ergodic theorem succeed or fail depending on the rate of convergence of l_n ? In particular is there a (slow) rate which still guarantees strong sweeping out? (Jones et al. 1999) contains some interesting conjectures in this direction.

There are also interesting questions concerning the mean and pointwise ergodic theorems for subsequences which are chosen randomly in some sense. See Bourgain (1988c; Jones et al. 1999) for some results in this direction. Again (Jones et al. 1999) contains some interesting conjectures along these lines.

In a recent paper (Bergelson and Leibman 2007) Bergelson and Leibman prove some very interesting and surprising results about the distribution of *generalized polynomials*. A generalized polynomial is any function which can be built starting with polynomials in $\mathbb{R}[x]$ using the operations of addition, multiplication and taking the greatest integer. As a consequence they derive a generalization of von Neumann's mean ergodic theorem to averages along generalized polynomial sequences. The following is a special case.

Theorem 51 *Suppose p is a generalized polynomial taking integer values on the integers and U is a unitary operator on $\mathcal{H}$. Then $\frac{1}{n} \sum_{i=0}^{n-1} U^{p(i)} x$ is norm convergent for all $x \in \mathcal{H}$.*

This begs the question: does one have pointwise convergence? If so, this would be a far-reaching generalization of Bourgain's polynomial ergodic theorem.

There are also lots of questions concerning the nature of Følner sequences $\{F_n\}$ in an amenable group which give a pointwise theorem. For example Lindenstrauss (1999) has shown that in the *lamplighter* group, a semi-direct product of $\mathbb{Z}$ with $\bigoplus_{i \in \mathbb{Z}} \mathbb{Z}/2\mathbb{Z}$ on which Z acts by the shift, there is no sequence satisfying Tempelman's condition and that any $\{F_n\}$ satisfying the Shulman condition must grow super-exponentially. So, it is

natural to ask for slower rates of growth. In particular, in any amenable group is there always a sequence $\{F_n\}$ which is pointwise good and grows at most exponentially? Can one do better either in general or in particular groups?

Lindenstrauss's theorem at least guarantees the existence of Følner sequences which are pointwise good in L_1 but in particular groups there are often natural sequences which one hopes might be good. For example in $\bigoplus_{i \in \mathbb{Z}} \mathbb{Z}$ one may take F_n to be a cube based at 0 of side length l_n and dimension d_n (that is, all but the first d_n co-ordinates are zero), where both sequences increase to ∞ . What conditions on l_n and d_n will give a good sequence? Note that no such sequence is regular in Tempelman's sense. If $d_n = n$ then $\{l_n\}$ must be super exponential to ensure Shulman's condition. Can one do better? What about $l_n = d_n = n$?

Bibliography

- Akcoglu MA (1975) A pointwise ergodic theorem in L_p -spaces. Can J Math 27(5):1075–1082
- Akcoglu MA, Chacon RV (1965) A convexity theorem for positive operators. Z Wahrsch Verw Gebiete 3:328–332
- Akcoglu MA, Chacon RV (1970) A local ratio theorem. Can J Math 22:545–552
- Akcoglu MA, del Junco A (1975) Convergence of averages of point transformations. Proc Amer Math Soc 49: 265–266
- Akcoglu MA, Kopp PE (1977) Construction of dilations of positive L_p -contractions. Math Z 155(2):119–127
- Akcoglu MA, Krengel U (1981) Ergodic theorems for superadditive processes. J Reine Angew Math 323: 53–67
- Akcoglu MA, Sucheston L (1978) A ratio ergodic theorem for superadditive processes. Z Wahrsch Verw Gebiete 44(4):269–278
- Akcoglu M, Bellow A, Jones RL, Losert V, Reinhold-Larsson K, Wierdl M (1996) The strong sweeping out property for lacunary sequences, Riemann sums, convolution powers, and related matters. Ergod Theory Dyn Sys 16(2):207–253
- Akcoglu M, Jones RL, Rosenblatt JM (2000) The worst sums in ergodic theory. Michigan Math J 47(2): 265–285
- Alaoglu L, Birkhoff G (1939) General ergodic theorems. Proc Nat Acad Sci USA 25:628–630
- Assani I (2000) Multiple return times theorems for weakly mixing systems. Ann Inst H Poincaré Probab Statist 36(2):153–165

- Assani I (2003) Wiener–Wintner ergodic theorems. World Scientific Publishing Co. Inc, River Edge
- Assani I, Lesigne E, Rudolph D (1995) Wiener–Wintner return-times ergodic theorem. *Israel J Math* 92(1–3):375–395
- Assani I, Buczolic Z, Mauldin RD (2005) An L^1 counting problem in ergodic theory. *J Anal Math* 95:221–241
- Auslander L, Green L, Hahn F (1963) Flows on homogeneous spaces. With the assistance of Markus L, Massey W and an appendix by Greenberg L *Annals of Mathematics Studies*, No 53. Princeton University Press, Princeton, NJ
- Banach S (1993) *Théorie des opérations linéaires*. Éditions Jacques Gabay, Sceaux, reprint of the 1932 original
- Bellow A (1983) On “bad universal” sequences in ergodic theory II. In: Belley JM, Dubois J, Morales P (eds) *Measure theory and its applications*. Lecture notes in math, vol 1033. Springer, Berlin, pp 74–78
- Bellow A (1999) Transference principles in ergodic theory. In: Christ M, Kenig CE, Sadowsky C (eds) *Harmonic analysis and partial differential equations*, Chicago lectures in math. University of Chicago Press, Chicago, pp 27–39
- Bellow A, Calderón A (1999) A weak-type inequality for convolution products. In: Christ M, Kenig CE, Sadowsky C (eds) *Harmonic analysis and partial differential equations*, Chicago lectures in math. University of Chicago Press, Chicago, pp 41–48
- Bellow A, Jones R (eds) (1991) *Almost everywhere convergence, II*. Academic Press, Boston
- Bellow A, Losert V (1985) The weighted pointwise ergodic theorem and the individual ergodic theorem along subsequences. *Trans Am Math Soc* 288(1):307–345
- Bellow A, Jones R, Rosenblatt J (1989) Almost everywhere convergence of powers. In: Edgar GA, Sucheston L (eds) *Almost everywhere convergence*. Academic, Boston, pp 99–120
- Bellow A, Jones R, Rosenblatt J (1990) Convergence for moving averages. *Ergod Theory Dyn Sys* 10(1):43–62
- Bellow A, Jones RL, Rosenblatt J (1992) Almost everywhere convergence of weighted averages. *Math Ann* 293(3):399–426
- Bellow A, Jones R, Rosenblatt J (1994) Almost everywhere convergence of convolution powers. *Ergod Theory Dyn Sys* 14(3):415–432
- Bergelson V (1987) Weakly mixing PET. *Ergod Theory Dyn Sys* 7(3):337–349
- Bergelson V (1996) Ergodic Ramsey theory—an update. In: Pollicott M, Schmidt K (eds) *Ergodic theory of $\mathbb{Z}^d$ actions*. London Math Soc Lecture Note Ser, vol 228. Cambridge University Press, Cambridge, pp 1–61
- Bergelson V (2006) Combinatorial and Diophantine applications of ergodic theory. In: Hasselblatt B, Katok A (eds) *Handbook of dynamical systems*, vol 1B, Appendix A by Leibman A, Appendix B by Quas A, Wierdl M. Elsevier, Amsterdam, pp 745–869
- Bergelson V (2007) Some historical remarks and modern questions around the ergodic theorem. *Int Math Nachr* 205:1–10
- Bergelson V, Leibman A (2002) A nilpotent Roth theorem. *Invent Math* 147(2):429–470
- Bergelson V, Leibman A (2004) Failure of the Roth theorem for solvable groups of exponential growth. *Ergod Theory Dyn Sys* 24(1):45–53
- Bergelson V, Leibman A (2007) Distribution of values of bounded generalized polynomials. *Acta Math* 198(2):155–230
- Berkson E, Bourgain J, Gillespie TA (1991) On the almost everywhere convergence of ergodic averages for power-bounded operators on L^p -subspaces. *Integral Equ Oper Theory* 14(5):678–715
- Billingsley P (1965) *Ergodic theory and information*. Wiley, New York
- Birkhoff GD (1931) Proof of the ergodic theorem. *Proc Natl Acad Sci U S A* 17:656–660
- Bishop E (1966) An upcrossing inequality with applications. *Michigan Math J* 13:1–13
- Bishop E (1967/1968) A constructive ergodic theorem. *J Math Mech* 17:631–639
- Blum JR, Hanson DL (1960) On the mean ergodic theorem for subsequences. *Bull Amer Math Soc* 66:308–311
- Bourgain J (1987) On pointwise ergodic theorems for arithmetic sets. *C R Acad Sci* 305(10):397–402
- Bourgain J (1988a) Almost sure convergence and bounded entropy. *Israel J Math* 63(1):79–97
- Bourgain J (1988b) An approach to pointwise ergodic theorems. In: Lindenstrauss J, Milman VD (eds) *Geometric aspects of functional analysis (1986/87)*, Lecture notes in math, vol 1317. Springer, Berlin, pp 204–223
- Bourgain J (1988c) On the maximal ergodic theorem for certain subsets of the integers. *Israel J Undergrad Math* 61(1):39–72
- Bourgain J (1988d) On the pointwise ergodic theorem on L^p for arithmetic sets. *Israel J Math* 61(1):73–84
- Bourgain J (1988e) Temps de retour pour les systèmes dynamiques. *C R Acad Sci* 306(12):483–485
- Bourgain J (1989a) Almost sure convergence in ergodic theory. In: Edgar GA, Sucheston L (eds) *Almost everywhere convergence*. Academic, Boston, pp 145–151
- Bourgain J (1989b) Pointwise ergodic theorems for arithmetic sets. *Inst Hautes Études Sci Publ Math* (69):5–45, with an appendix by the author, Furstenberg H, Katznelson Y, Ornstein DS
- Bourgain J (1990) Double recurrence and almost sure convergence. *J Reine Angew Math* 404:140–161
- Breiman L (1957) The individual ergodic theorem of information theory. *Ann Math Statist* 28:809–811
- Breiman L (1968) *Probability*. Addison-Wesley, Reading
- Brunel A, Keane M (1969) Ergodic theorems for operator sequences. *Z Wahrsch Verw Gebiete* 12:231–240
- Burkholder DL (1962) Semi-Gaussian subspaces. *Trans Am Math Soc* 104:123–131
- Calderon AP (1953) A general ergodic theorem. *Ann Math* 2(58):182–191
- Calderón AP (1968) Ergodic theory and translation-invariant operators. *Proc Natl Acad Sci U S A* 59:349–353

- Calderon AP, Zygmund A (1952) On the existence of certain singular integrals. *Acta Math* 88:85–139
- Chacon RV (1962) Identification of the limit of operator averages. *J Math Mech* 11:961–968
- Chacon RV (1963a) Convergence of operator averages. In: Wright FB (ed) *Ergodic theory*. Academic, New York, pp 89–120
- Chacon RV (1963b) Linear operators in L_1 . In: Wright FB (ed) *Ergodic theory*. Academic, New York, pp 75–87
- Chacon RV (1964) A class of linear transformations. *Proc Am Math Soc* 15:560–564
- Chacon RV, Ornstein DS (1960) A general ergodic theorem. *Ill J Math* 4:153–160
- Conze JP, Lesigne E (1984) Théorèmes ergodiques pour des mesures diagonales. *Bull Soc Math France* 112(2): 143–175
- Conze JP, Lesigne E (1988) Sur un théorème ergodique pour des mesures diagonales. In: *Probabilités*, Publ Inst Rech Math Rennes, vol 1987. University Rennes I, Rennes, pp 1–31
- Cotlar M (1955a) On ergodic theorems. *Math Notae* 14: 85–119. (1956)
- Cotlar M (1955b) A unified theory of Hilbert transforms and ergodic theorems. *Rev Mat Cuyana* 1:105–167. (1956)
- Day M (1942) Ergodic theorems for Abelian semigroups. *Trans Am Math Soc* 51:399–412
- del Junco A, Rosenblatt J (1979) Counterexamples in ergodic theory and number theory. *Math Ann* 245(3): 185–197
- Demeter C, Lacey M, Tao T, Thiele C (2008) Breaking the duality in the return times theorem. *Duke Math J* 143(2):281–355
- Derrien JM, Lesigne E (1996) Un théorème ergodique polynomial ponctuel pour les endomorphismes exacts et les K-systèmes. *Ann Inst H Poincaré Probab Statist* 32(6):765–778
- Derriennic Y (2006) Some aspects of recent works on limit theorems in ergodic theory with special emphasis on the “central limit theorem”. *Discrete Contin Dyn Syst* 15(1):143–158
- Doob JL (1938) Stochastic processes with an integral-valued parameter. *Trans Amer Math Soc* 44(1):87–150
- Dunford N (1951) An individual ergodic theorem for non-commutative transformations. *Acta Sci Math Szeged* 14:1–4
- Dunford N, Schwartz J (1955) Convergence almost everywhere of operator averages. *Proc Natl Acad Sci U S A* 41:229–231
- Dunford N, Schwartz JT (1988) *Linear operators, part I*. Wiley Classics Library, Wiley, New York. General theory, with the assistance of Bade WG, Bartle RG, Reprint of the 1958 original, A Wiley–Interscience Publication
- Durand S, Schneider D (2003) Random ergodic theorems and regularizing random weights. *Ergod Theory Dyn Sys* 23(4):1059–1092
- Eberlein WF (1949) Abstract ergodic theorems and weak almost periodic functions. *Trans Am Math Soc* 67: 217–240
- Edgar G, Sucheston L (eds) (1989) *Almost everywhere convergence*. Academic Press, Boston
- Emerson WR (1974) The pointwise ergodic theorem for amenable groups. *Am J Math* 96:472–487
- Feldman J (2007) A ratio ergodic theorem for commuting, conservative, invertible transformations with quasi-invariant measure summed over symmetric hypercubes. *Ergod Theory Dyn Sys* 27(4):1135–1142
- Foguel SR (1969) The ergodic theory of Markov processes. In: *Van Nostrand Mathematical Studies*, No 21. Van Nostrand Reinhold, New York
- Frantzikinakis N, Kra B (2005a) Convergence of multiple ergodic averages for some commuting transformations. *Ergod Theory Dyn Sys* 25(3):799–809
- Frantzikinakis N, Kra B (2005b) Polynomial averages converge to the product of integrals. *Israel J Math* 148:267–276
- Frantzikinakis N, Kra B (2006) Ergodic averages for independent polynomials and applications. *J London Math Soc* 2) 74(1):131–142
- Furstenberg H (1967) Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math Sys Theory* 1:1–49
- Furstenberg H (1977) Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J Analyse Math* 31:204–256
- Furstenberg H (1981) *Recurrence in ergodic theory and combinatorial number theory*. Princeton University Press, Princeton, m B Porter Lectures
- Furstenberg H, Kesten H (1960) Products of random matrices. *Ann Math Statist* 31:457–469
- Furstenberg H, Weiss B (1996) A mean ergodic theorem for $(1/N) \sum_{n=1}^N f(T^n x)g(T^{n^2} x)$. In: Bergelson V, March-P, Rosenblatt J (eds) *Convergence in ergodic theory and probability*, Ohio State Univ Math Res Inst Publ, vol 5. de Gruyter, Berlin, pp 193–227
- Garsia AM (1970) Topics in almost everywhere convergence. In: *Lectures in advanced mathematics*, vol 4. Markham, Chicago
- Glasner E (2003) *Ergodic theory via joinings*, Mathematical surveys and monographs, vol 101. American Mathematical Society, Providence
- Gowers WT (2001) A new proof of Szemerédi’s theorem. *Geom Funct Anal* 11(3):465–588
- Green B, Tao T (2004) The primes contain arbitrarily large arithmetic progressions. <http://arxiv.org/abs/mathNT/0404188>
- Greenleaf FP (1973) Ergodic theorems and the construction of summing sequences in amenable locally compact groups. *Commun Pure Appl Math* 26:29–46
- Greenleaf FP, Emerson WR (1974) Group structure and the pointwise ergodic theorem for connected amenable groups. *Adv Math* 14:153–172

- Guivarc'h Y (1969) Généralisation d'un théorème de von Neumann. *C R Acad Sci Paris Sér A-B* 268: A1020–A1023
- Halmos PR (1946) An ergodic theorem. *Proc Natl Acad Sci U S A* 32:156–161
- Halmos PR (1949) A nonhomogeneous ergodic theorem. *Trans Am Math Soc* 66:284–288
- Halmos PR (1960) *Lectures on ergodic theory*. Chelsea, New York
- Hammersley JM, Welsh DJA (1965) First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. In: *Proc Internat Res Semin. Statist Lab, University of California, Berkeley*. Springer, New York, pp 61–110
- Hopf E (1937) *Ergodentheorie*. In: *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol 5. Springer, Berlin
- Hopf E (1954) The general temporally discrete Markoff process. *J Ration Mech Anal* 3:13–45
- Host B, Kra B (2001) Convergence of Conze–Lesigne averages. *Ergod Theory Dyn Sys* 21(2):493–509
- Host B, Kra B (2005a) Convergence of polynomial ergodic averages. *Israel J Math* 149:1–19
- Host B, Kra B (2005b) Nonconventional ergodic averages and nilmanifolds. *Ann Math* 161(1):397–488
- Hurewicz W (1944) Ergodic theorem without invariant measure. *Ann Math* 2(45):192–206
- Ivanov VV (1996a) Geometric properties of monotone functions and the probabilities of random oscillations. *Sibirsk Mat Zh* 37(1):117–115
- Ivanov VV (1996b) Oscillations of averages in the ergodic theorem. *Dokl Akad Nauk* 347(6):736–738
- Jewett RI (1969/1970) The prevalence of uniquely ergodic systems. *J Math Mech* 19:717–729
- Jones RL (1987) Necessary and sufficient conditions for a maximal ergodic theorem along subsequences. *Ergod Theory Dyn Sys* 7(2):203–210
- Jones RL (1993) Ergodic averages on spheres. *J Anal Math* 61:29–45
- Jones RL, Wierdl M (1994) Convergence and divergence of ergodic averages. *Ergod Theory Dyn Sys* 14(3): 515–535
- Jones RL, Olsen J, Wierdl M (1992) Subsequence ergodic theorems for L^p contractions. *Trans Am Math Soc* 331(2):837–850
- Jones R, Rosenblatt J, Tempelman A (1994) Ergodic theorems for convolutions of a measure on a group. *Ill J Math* 38(4):521–553
- Jones RL, Kaufman R, Rosenblatt JM, Wierdl M (1998) Oscillation in ergodic theory. *Ergod Theory Dyn Sys* 18(4):889–935
- Jones RL, Lacey M, Wierdl M (1999) Integer sequences with big gaps and the pointwise ergodic theorem. *Ergod Theory Dyn Sys* 19(5):1295–1308
- Jones RL, Rosenblatt JM, Wierdl M (2001) Oscillation inequalities for rectangles. *Proc Am Math Soc* 129(5): 1349–1358. (electronic)
- Jones RL, Wierdl M (2003) Oscillation in ergodic theory: higher dimensional results. *Israel J Math* 135: 1–27
- Kac M (1947) On the notion of recurrence in discrete stochastic processes. *Bull Amer Math Soc* 53: 1002–1010
- Kachurovskii AG (1996a) Rates of convergence in ergodic theorems. *Uspekhi Mat Nauk* 51(4(310)):73–124
- Kachurovskii AG (1996b) Spectral measures and convergence rates in the ergodic theorem. *Dokl Akad Nauk* 347(5):593–596
- Kakutani S (1940) Ergodic theorems and the Markoff process with a stable distribution. *Proc Imp Acad Tokyo* 16:49–54
- Kalikow S, Weiss B (1999) Fluctuations of ergodic averages. In: *Proceedings of the conference on probability, ergodic theory, and analysis*, vol 43, pp 480–488
- Kamae T (1982) A simple proof of the ergodic theorem using nonstandard analysis. *Israel J Math* 42(4): 284–290
- Katok A, Hasselblatt B (1995) *Introduction to the modern theory of dynamical systems*, *Encyclopedia of Mathematics and its Applications*, vol 54. Cambridge University Press, Cambridge, with a supplementary chapter by Katok A and Mendoza L
- Katznelson Y, Weiss B (1982) A simple proof of some ergodic theorems. *Israel J Math* 42(4):291–296
- Kieffer JC (1975) A generalized Shannon–McMillan theorem for the action of an amenable group on a probability space. *Ann Probab* 3(6):1031–1037
- Kingman JFC (1968) The ergodic theory of subadditive stochastic processes. *J Roy Stat Soc Ser B* 30:499–510
- Kingman JFC (1976) Subadditive processes. In: *École d'Été de Probabilités de Saint-Flour, V–1975*, *Lecture notes in math*, vol 539. Springer, Berlin, pp 167–223
- Koopman B (1931) Hamiltonian systems and transformations in Hilbert spaces. *Proc Natl Acad Sci U S A* 17: 315–318
- Krantz SG, Parks HR (1999) *The geometry of domains in space*. Birkhäuser Advanced Texts: Basler Lehrbücher. Birkhäuser Boston Inc, Boston
- Krengel U (1971) On the individual ergodic theorem for subsequences. *Ann Math Stat* 42:1091–1095
- Krengel U (1978/79) On the speed of convergence in the ergodic theorem. *Monatsh Math* 86(1):3–6
- Krengel U (1985) Ergodic theorems. In: *de Gruyter studies in mathematics*, vol 6. de Gruyter, Berlin. with a supplement by Antoine Brunel
- Krengel U, Lin M, Wittmann R (1990) A limit theorem for order preserving nonexpansive operators in L^1 . *Israel J Undergrad Math* 71(2):181–191
- Krieger W (1972) On unique ergodicity. In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol II. Probability Theory, University of California Press, Berkeley, pp. 327–346

- Kryloff N, Bogoliouboff N (1937) La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire. *Ann Math* 38(1):65–113
- Kuipers L, Niederreiter H (1974) Uniform distribution of sequences. Wiley, New York, Pure and Applied Mathematics
- Lamperti J (1958) On the isometries of certain function-spaces. *Pac J Math* 8:459–466
- Leibman A (2005a) Convergence of multiple ergodic averages along polynomials of several variables. *Israel J Math* 146:303–315
- Leibman A (2005b) Pointwise convergence of ergodic averages for polynomial actions of $\mathbb{Z}^d$ by translations on a nilmanifold. *Ergod Theory Dyn Sys* 25(1): 215–225
- Lemańczyk M, Lesigne E, Parreau F, Volný D, Wierdl M. n.d.
- Lesigne E (1989) Théorèmes ergodiques pour une translation sur un nilvariété. *Ergod Theory Dyn Sys* 9(1): 115–126
- Lindenstrauss E (1999) Pointwise theorems for amenable groups. *Electron Res Announc Amer Math Soc* 5: 82–90
- Loomis LH (1946) A note on the Hilbert transform. *Bull Amer Math Soc* 52:1082–1086
- Lorch ER (1939) Means of iterated transformations in reflexive vector spaces. *Bull Amer Math Soc* 45: 945–947
- Mauldin D, Buczolich Z (2005) Concepts behind divergent ergodic averages along the squares. In: Assani I (ed) *Ergodic theory and related fields*, *Contemp math*, vol 430. American Mathematical Society, Providence, pp 41–56
- McMillan B (1953) The basic theorems of information theory. *Ann Math Stat* 24:196–219
- Merlevède F, Peligrad M, Utev S (2006) Recent advances in invariance principles for stationary sequences. *Probab Surv* 3:1–36
- Neveu J (1961) Sur le théorème ergodique ponctuel. *C R Acad Sci Paris* 252:1554–1556
- Neveu J (1965) Mathematical foundations of the calculus of probability. Trans. Amiel Feinstein. Holden-Day, San Francisco
- Nevo A (1994) Harmonic analysis and pointwise ergodic theorems for noncommuting transformations. *J Am Math Soc* 7(4):875–902
- Nevo A (2006) Pointwise ergodic theorems for actions of groups. In: Hasselblatt B, Katok A (eds) *Handbook of dynamical systems*, vol 1B. Elsevier, Amsterdam, pp 871–982
- Nevo A, Stein EM (1994) A generalization of Birkhoff's pointwise ergodic theorem. *Acta Math* 173(1):135–154
- Nguyen XX (1979) Ergodic theorems for subadditive spatial processes. *Z Wahrsch Verw Gebiete* 48(2):159–176
- Orey S (1971) Lecture notes on limit theorems for Markov chain transition probabilities. In: Van Nostrand Reinhold mathematical studies, no 34. Van Nostrand Reinhold Co., London
- Ornstein DS (1960) On invariant measures. *Bull Amer Math Soc* 66:297–300
- Ornstein D (1970) Bernoulli shifts with the same entropy are isomorphic. *Adv Math* 4:337–352
- Ornstein D (1971) A remark on the Birkhoff ergodic theorem. *Ill J Math* 15:77–79
- Ornstein D, Weiss B (1983) The Shannon–McMillan–Breiman theorem for a class of amenable groups. *Israel J Math* 44(1):53–60
- Ornstein D, Weiss B (1992) Subsequence ergodic theorems for amenable groups. *Israel J Math* 79(1):113–127
- Oseledec VI (1968) A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems. *Trudy Moskov Mat Obšč* 19:179–210
- Oxtoby JC (1952) Ergodic sets. *Bull Amer Math Soc* 58: 116–136
- Parry W (1969) Ergodic properties of affine transformations and flows on nilmanifolds. *Am J Math* 91:757–771
- Paterson A (1988) Amenability, mathematical surveys and monographs, vol 29. American Mathematical Society, Providence
- Peck JEL (1951) An ergodic theorem for a non-commutative semigroup of linear operators. *Proc Am Math Soc* 2:414–421
- Pesin JB (1977) Characteristic Ljapunov exponents, and smooth ergodic theory. *Uspehi Mat Nauk* 32(4(196)):55–112,287
- Petersen K (1983) Another proof of the existence of the ergodic Hilbert transform. *Proc Am Math Soc* 88(1): 39–43
- Petersen K (1989) *Ergodic theory*, Cambridge studies in advanced mathematics, vol 2. Cambridge University Press, Cambridge
- Pitt HR (1942) Some generalizations of the ergodic theorem. *Proc Camb Philos Soc* 38:325–343
- Poincaré H (1987) Les méthodes nouvelles de la mécanique céleste. Tome I, II, III. Les Grands Classiques Gauthier–Villars. Librairie Scientifique et Technique Albert Blanchard, Paris
- Raghunathan MS (1979) A proof of Oseledec's multiplicative ergodic theorem. *Israel J Math* 32(4):356–362
- Renaud PF (1971) General ergodic theorems for locally compact groups. *Am J Math* 93:52–64
- Rosenblatt JM, Wierdl M (1995) Pointwise ergodic theorems via harmonic analysis. In: Peterson KE, Salama IA (eds) *Ergodic theory and its connections with harmonic analysis*, London Mathematical Society lecture note series, vol 205. Cambridge University Press, Cambridge, pp 3–151
- Rudolph DJ (1990) Fundamentals of measurable dynamics. In: *Fundamentals of measurable dynamics: ergodic theory on Lebesgue spaces*. Oxford Science Publications, The Clarendon Press, Oxford University Press, New York
- Rudolph DJ (1994) A joinings proof of Bourgain's return time theorem. *Ergod Theory Dyn Sys* 14(1):197–203

- Rudolph DJ (1998) Fully generic sequences and a multiple-term return-times theorem. *Invent Math* 131(1):199–228
- Ruelle D (1982) Characteristic exponents and invariant manifolds in Hilbert space. *Ann Math* 115(2):243–290
- Ryll-Nardzewski C (1951) On the ergodic theorems, II. Ergodic theory of continued fractions. *Stud Math* 12:74–79
- Ryzhikov V (1994) Joinings, intertwining operators, factors and mixing properties of dynamical systems. *Russian Acad Sci Izv Math* 42:91–114
- Shah NA (1998) Invariant measures and orbit closures on homogeneous spaces for actions of subgroups generated by unipotent elements. In: Dani SG (ed) *Lie groups and ergodic theory*. Tata Inst Fund Res Stud Math, vol 14. Tata Institute of Fundamental Research, Bombay, pp 229–271
- Shalom Y (1998) Random ergodic theorems, invariant means and unitary representation. In: Dani SG (ed) *Lie groups and ergodic theory*. Tata Inst Fund Res Stud Math, vol 14. Tata Institute of Fundamental Research, Bombay, pp 273–314
- Shannon CE (1948) A mathematical theory of communication. *Bell System Tech J* 27(379–423):623–656
- Shields PC (1987) The ergodic and entropy theorems revisited. *IEEE Trans Inform Theory* 33(2):263–266
- Shulman A (1988) Maximal ergodic theorems on groups. *Dep Lit NIINTI* No. 2184
- Sine R (1970) A mean ergodic theorem. *Proc Am Math Soc* 24:438–439
- Smythe RT (1976) Multiparameter subadditive processes. *Ann Probab* 4(5):772–782
- Szemerédi E (1975) On sets of integers containing no k elements in arithmetic progression. *Acta Arith* 27:199–245
- Tao T (2005) The Gaussian primes contain arbitrarily shaped constellations. <http://arxiv.org/abs/math/0501314>
- Tao T (2008) Norm convergence of multiple ergodic averages for commuting transformations. *Ergod Theory Dyn Syst* 28(2):657–688
- Tao T, Ziegler T (2006) The primes contain arbitrarily long polynomial progressions. <http://frontmath.ucdavis.edu/06105050>
- Tempelman A (1992) *Ergodic theorems for group actions, mathematics and its applications*, vol 78. Kluwer, Dordrecht, informational and thermodynamical aspects, Translated and revised from the 1986 Russian original
- Tempel'man AA (1967) Ergodic theorems for general dynamical systems. *Dokl Akad Nauk SSSR* 176:790–793
- Tempel'man AA (1972a) Ergodic theorems for general dynamical systems. *Trudy Moskov Mat Obšč* 26:95–132
- Tempel'man AA (1972b) A generalization of a certain ergodic theorem of Hopf. *Teor Veroyatnost i Primenen* 17:380–383
- Thouvenot JP (1995) Some properties and applications of joinings in ergodic theory. In: Peterson KE, Salama IA (eds) *Ergodic theory and its connections with harmonic analysis*, London Math Soc Lecture Note Ser, vol 205. Cambridge University Press, Cambridge, pp 207–235
- Tulcea AI (1964) Ergodic properties of positive isometries. *Bull AMS* 70:366–371
- von Neumann J (1932) Proof of the quasi-ergodic hypothesis. *Proc Natl Acad Sci USA* 18:70–82
- Walters P (1993) A dynamical proof of the multiplicative ergodic theorem. *Trans Am Math Soc* 335(1):245–257
- Weber M (1998) Entropie métrique et convergence presque partout, *Travaux en Cours*, vol 58. Hermann, Paris
- Weiss B (2003) Actions of amenable groups. In: Bezuglyi S, Kolyada S (eds) *Topics in dynamics and ergodic theory*, London Math Soc Lecture Note Ser, vol 310. Cambridge University Press, Cambridge, pp 226–262
- Weyl H (1916) Über die Gleichverteilung von Zahlen mod Eins. *Math Ann* 77(3):313–352
- Wiener N (1939) The ergodic theorem. *Duke Math J* 5(1):1–18
- Wiener N, Wintner A (1941) Harmonic analysis and ergodic theory. *Am J Math* 63:415–426
- Wierdl M (1988) Pointwise ergodic theorem along the prime numbers. *Israel J Math* 64(3):315–336. (1989)
- Wittmann R (1995) Almost everywhere convergence of ergodic averages of nonlinear operators. *J Funct Anal* 127(2):326–362
- Yosida K (1940a) An abstract treatment of the individual ergodic theorem. *Proc Imp Acad Tokyo* 16:280–284
- Yosida K (1940b) Ergodic theorems of Birkhoff–Khinchine's type. *Jap J Math* 17:31–36
- Yosida K, Kakutani S (1939) Birkhoff's ergodic theorem and the maximal ergodic theorem. *Proc Imp Acad, Tokyo* 15:165–168
- Ziegler T (2005) A non-conventional ergodic theorem for a nilsystem. *Ergod Theory Dyn Sys* 25(4):1357–1370
- Ziegler T (2007) Universal characteristic factors and Furstenberg averages. *J Am Math Soc* 20(1):53–97. (electronic)
- Zimmer RJ (1976a) Ergodic actions with generalized discrete spectrum. *Ill J Math* 20(4):555–588
- Zimmer RJ (1976b) Extensions of ergodic group actions. *Ill J Math* 20(3):373–409
- Zund JD (2002) George David Birkhoff and John von Neumann: a question of priority and the ergodic theorems, 1931–1932. *Hist Math* 29(2):138–156
- Zygmund A (1951) An individual ergodic theorem for non-commutative transformations. *Acta Sci Math Szeged* 14:103–110

Spectral Theory of Dynamical Systems

Adam Kanigowski¹ and Mariusz Lemańczyk²

¹Department of Mathematics, University of Maryland at College Park, College Park, MD, USA

²Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland

Article Outline

Glossary and Notation

Definition of the Subject

Introduction

Maximal Spectral Type of a Koopman

Representation: Alexeyev's Theorem

Spectral Theory of Weighted Operators

The Multiplicity Function

Rokhlin Cocycles

Rank-1 and Related Systems

Spectral Theory of Dynamical Systems of

Probabilistic Origin

Inducing and Spectral Theory

Rigid Sequences

Spectral Theory of Parabolic Dynamical Systems

Spectral Theory for Locally Compact Groups of

Type I

Future Directions

Bibliography

Glossary and Notation

AT property of an automorphism An automorphism T of a standard probability Borel space $(X, \mathcal{B}, \mu)$ is called *approximatively transitive* (AT for short) if for every $\varepsilon > 0$ and every finite set $f_1, \dots, f_n$ of non-negative L^1 -functions on $(X, \mathcal{B}, \mu)$ we can find $f \in L^1(X, \mathcal{B}, \mu)$ also non-negative such that $\|f_i - \sum_j \alpha_{ij} f \circ T^{n_j}\|_{L^1} < \varepsilon$ for all $i = 1, \dots, n$ (for some $\alpha_{ij} \geq 0, n_j \in \mathbb{N}$).

Ergodicity, weak mixing, mild mixing, mixing, and rigidity of $\mathcal{T}$ A measure-

preserving $\mathbb{A}$ -action $\mathcal{T} = (T_a)_{a \in \mathbb{A}}$ is called *ergodic* if $\chi_0 \equiv 1 \in \hat{\mathbb{A}}$ is a simple eigenvalue of $\mathcal{U}_{\mathcal{T}}$. It is *weakly mixing* if $\mathcal{U}_{\mathcal{T}}$ has a continuous spectrum on the subspace $L_0^2(X, \mathcal{B}, \mu)$ of zero mean functions. $\mathcal{T}$ is said to be *rigid* if there is a sequence (a_n) going to infinity in $\mathbb{A}$ such that the sequence $(U_{T_{a_n}})$ goes to the identity in the strong (or weak) operator topology; $\mathcal{T}$ is said to be *mildly mixing* if it has no non-trivial rigid factors. We say that $\mathcal{T}$ is *mixing* if the operator equal to zero is the only limit point of $\{U_{T_a}|_{L_0^2(X, \mathcal{B}, \mu)} : a \in \mathbb{A}\}$ in the weak operator topology.

Cocycles and group extensions If T is an ergodic automorphism, G is an l.c.s.c. Abelian group, and $\varphi : X \rightarrow G$ is measurable, then the pair (T, φ) generates a *cocycle* $\varphi^{(\cdot)}(\cdot) : \mathbb{Z} \times X \rightarrow G$, where $\varphi^{(n)}(x) =$

$$\begin{cases} \varphi(x) + \dots + \varphi(T^{n-1}x) & \text{for } n > 0, \\ 0 & \text{for } n = 0 \\ -(\varphi(T^n x) + \dots + \varphi(T^{-1}x)) & \text{for } n < 0. \end{cases}$$

(That is, $(\varphi^{(n)})$ is a standard 1-cocycle in the algebraic sense for the $\mathbb{Z}$ -action $n(f) = f \circ T^n$ on the group of measurable functions on X with values in G ; hence, the function $\varphi : X \rightarrow G$ itself is often called a *cocycle*.) Assume additionally that G is compact. Using the cocycle φ , we define a *group extension* T_φ on $(X \times G, \mathcal{B} \otimes \mathcal{B}(G), \mu \otimes \lambda_G)$ (λ_G stands for Haar measure of G), where $T_\varphi(x, g) = (Tx, \varphi(x) + g)$.

Induced automorphism Assume that T is an automorphism of a standard probability Borel space $(X, \mathcal{B}, \mu)$. Let $A \in \mathcal{B}$, $\mu(A) > 0$. The *induced automorphism* T_A is defined on the conditional space $(X, \mathcal{B}_A, \mu_A)$, where $\mathcal{B}_A$ is the trace of $\mathcal{B}$ on A , $\mu_A(B) = \mu(B)/\mu(A)$ for $B \in \mathcal{B}_A$ and $T_A(x) = T^{k_A(x)}x$, where $k_A(x)$ is the smallest $k \geq 1$ for which $T^k x \in A$.

Kolmogorov group property An $\mathbb{A}$ -action $\mathcal{T}$ satisfies the *Kolmogorov group property* if $\sigma_{\mathcal{U}_{\mathcal{T}}} * \sigma_{\mathcal{U}_{\mathcal{T}}} \ll \sigma_{\mathcal{U}_{\mathcal{T}}}$.

Koopman representation of a dynamical system $\mathcal{T}$ Let $\mathbb{A}$ be an l.c.s.c. (and not compact) Abelian group and $\mathcal{T} : a \mapsto T_a$ a representation of $\mathbb{A}$ in the group $\text{Aut}(X, \mathcal{B}, \mu)$ of (measure-preserving) automorphisms of a standard probability Borel space $(X, \mathcal{B}, \mu)$. The *Koopman representation* $\mathcal{U} = \mathcal{U}_{\mathcal{T}}$ of $\mathcal{T}$ in $L^2(X, \mathcal{B}, \mu)$ is defined as the unitary representation $a \mapsto U_{T_a} \in U(L^2(X, \mathcal{B}, \mu))$, where $U_{T_a}(f) = f \circ T_a$.

Markov operator A linear operator $J : L^2(X, \mathcal{B}, \mu) \rightarrow L^2(Y, \mathcal{C}, \nu)$ is called *Markov* if it sends non-negative functions to non-negative functions and $J1 = J^*1 = 1$.

Maximal spectral type and the multiplicity function of $\mathcal{U}$ The *maximal spectral type* $\sigma_{\mathcal{U}}$ of $\mathcal{U}$ is the type of σ_{x_1} in any spectral decomposition of H ; the *multiplicity function* $M_{\mathcal{U}} : \hat{\mathbb{A}} \rightarrow \{1, 2, \dots\} \cup \{+\infty\}$ is defined $\sigma_{\mathcal{U}}$ -a.e. and $M_{\mathcal{U}}(\chi) = \sum_{i=1}^{\infty} 1_{Y_i}(\chi)$, where $Y_1 = \hat{\mathbb{A}}$ and $Y_i = \text{supp} \frac{d\sigma_{x_i}}{d\sigma_{x_1}}$ for $i \geq 2$. A representation $\mathcal{U}$ is said to have *simple spectrum* if H is reduced to a single cyclic space. The multiplicity is *uniform* if there is only one essential value of $M_{\mathcal{U}}$. The essential supremum of $M_{\mathcal{U}}$ is called the *maximal spectral multiplicity*. $\mathcal{U}$ is said to have *discrete spectrum* if H has an orthonormal basis consisting of eigenvectors of $\mathcal{U}$; $\mathcal{U}$ has *singular (Haar, absolutely continuous) spectrum* if the maximal spectral type of $\mathcal{U}$ is singular with respect to (equivalent to, absolutely continuous with) a Haar measure of $\hat{\mathbb{A}}$.

SCS property We say that a Borel measure σ on $\hat{\mathbb{A}}$ satisfies the *strong convolution singularity property* (SCS property) if, for each $n \geq 1$, in the disintegration (given by the map $(\chi_1, \dots, \chi_n) \mapsto \chi_1 \cdot \dots \cdot \chi_n$) $\sigma^{\otimes n} = \int_{\hat{\mathbb{A}}} v_{\chi} d\sigma^{(n)}(\chi)$ the conditional measures v_{χ} are atomic with exactly $n!$ atoms ($\sigma^{(n)}$ stands for the n -th convolution $\sigma * \dots * \sigma$). An $\mathbb{A}$ -action $\mathcal{T}$ satisfies the SCS property if the maximal spectral type of $\mathcal{U}_{\mathcal{T}}$ on L^2_0 is a type of an SCS measure.

Special flow Given an ergodic automorphism T on a standard probability Borel space $(X, \mathcal{B}, \mu)$ and a positive integrable function $f : X \rightarrow \mathbb{R}^+$, we put $X^f = \{(x, t) \in X \times \mathbb{R} : 0 \leq t < f(x)\}$, $\mathcal{B}^f = \mathcal{B} \otimes$

$\mathcal{B}(\mathbb{R})|_{X^f}$, and define μ^f as normalized $\mu \otimes \lambda_{\mathbb{R}}|_{X^f}$. By a *special flow*, we mean the $\mathbb{R}$ -action $T^f = (T^f_t)_{t \in \mathbb{R}}$ under which a point $(x, s) \in X^f$ moves vertically with the unit speed, and once it reaches the graph of f , it is identified with $(Tx, 0)$.

Spectral decomposition of a unitary representation If $\mathcal{U} = (U_a)_{a \in \mathbb{A}}$ is a continuous unitary representation of a locally compact second countable (l.c.s.c.) Abelian group $\mathbb{A}$ in a separable Hilbert space H , then a decomposition $H = \bigoplus_{i=1}^{\infty} \mathbb{A}(x_i)$ is called *spectral* if $\sigma_{x_1} \gg \sigma_{x_2} \gg \dots$ (such a sequence of measures is also called *spectral*); here $\mathbb{A}(x) := \overline{\text{span}}\{U_a x : a \in \mathbb{A}\}$ is called the *cyclic space* generated by $x \in H$, and σ_x stands for the spectral measure of x .

Spectral disjointness Two $\mathbb{A}$ -actions $\mathcal{S}$ and $\mathcal{T}$ are called *spectrally disjoint* if the maximal spectral types of their Koopman representations $\mathcal{U}_{\mathcal{T}}$ and $\mathcal{U}_{\mathcal{S}}$ on the corresponding L^2_0 -spaces are mutually singular.

Time change Let $\mathcal{R} = (R_t)_{t \in \mathbb{R}}$ be a flow on $(X, \mathcal{B}, \mu)$ and let $v \in L^1(X, \mathcal{B}, \mu)$ be a positive function. The function v determines a cocycle over $\mathcal{R}$ given by the formula $v(t, x) := \int_0^t v(R_s x) ds$. Then for a.e. $x \in X$ and all $t \in \mathbb{R}$, there exists a unique $u = u(t, x)$ such that $\int_0^u v(R_s x) ds = t$. Now, we can define the flow $\tilde{\mathcal{R}}_t(x) := R_{u(t, x)}(x)$. The new flow $\tilde{\mathcal{R}} = (\tilde{R}_t)_{t \in \mathbb{R}}$ has the same orbits as the original flow, and it preserves the measure $\tilde{\mu} \ll \mu$ (hence, it is ergodic if $\mathcal{R}$ was), where $\frac{d\tilde{\mu}}{d\mu} = v / \int_X v d\mu$.

Unitary actions on Fock spaces If H is a separable Hilbert space, then by $H^{\odot n}$, we denote the subspace of n -tensors of $H^{\otimes n}$ symmetric under all permutations of coordinates, $n \geq 1$; then the Hilbert space $F(H) := \bigoplus_{n=0}^{\infty} H^{\odot n}$ is called a *symmetric Fock space*. If $V \in U(H)$, then $F(V) := \bigoplus_{n=0}^{\infty} V^{\odot n} \in U(F(H))$, where $V^{\odot n} = V^{\otimes n}|_{H^{\odot n}}$.

Weighted operator Let T be an ergodic automorphism of $(X, \mathcal{B}, \mu)$ and $\xi : X \rightarrow \mathbb{T}$ be a measurable function. The (unitary) operator $V = V_{\xi, T}$

acting on $L^2(X, \mathcal{B}, \mu)$ by the formula $V_{\xi T}(f)(x) = \xi(x)f(Tx)$ is called a *weighted operator*.

Definition of the Subject

Spectral theory of dynamical systems is a study of special unitary representations, called Koopman representations (see Section “Glossary and Notation”). Invariants of such representations are called spectral invariants of measure-preserving systems. Together with the entropy, they constitute the most important invariants used in the study of measure-theoretic intrinsic properties and classification problems of dynamical systems as well as in applications. Spectral theory was originated by von Neumann, Halmos, and Koopman in the 1930s. In this article, we will focus on recent progresses in the spectral theory of finite measure-preserving dynamical systems.

Introduction

Throughout $\mathbb{A}$ denotes a non-compact l.c.s.c. Abelian group ($\mathbb{A}$ will be most often $\mathbb{Z}$ or $\mathbb{R}$). The assumption of second countability implies that $\mathbb{A}$ is metrizable, and σ -compact and the space $C_0(\mathbb{A})$ are separable. Moreover, the dual group $\widehat{\mathbb{A}}$ is also l.c.s.c. Abelian.

General Unitary Representations

We are interested in *unitary*, that is, with values in the unitary group $U(H)$ of a Hilbert space H , (weakly) continuous representations $V : \mathbb{A} \ni a \mapsto V_a \in U(H)$ of such groups (the scalar valued maps $a \mapsto \langle V_a x, y \rangle$ are continuous for each $x, y \in H$).

Let $H = L^2(\widehat{\mathbb{A}}, \mathcal{B}(\widehat{\mathbb{A}}), \mu)$, where $\mathcal{B}(\widehat{\mathbb{A}})$ stands for the σ -algebra of Borel sets of $\widehat{\mathbb{A}}$ and $\mu \in M^+(\widehat{\mathbb{A}})$ (whenever X is an l.c.s.c. space, by $M(X)$, we denote the set of complex Borel measures on X , while $M^+(X)$ stands for the subset of positive (finite) measures). Given $a \in \mathbb{A}$, for $f \in L^2(\widehat{\mathbb{A}}, \mathcal{B}(\widehat{\mathbb{A}}), \mu)$ put

$$V_a^\mu(f)(\chi) = i(a)(\chi) \cdot f(\chi) = \chi(a) \cdot f(\chi) \quad (\chi \in \widehat{\mathbb{A}}),$$

where $i : \mathbb{A} \rightarrow \widehat{\widehat{\mathbb{A}}}$ is the canonical Pontriagin isomorphism of $\mathbb{A}$ with its second dual. Then $V^\mu = (V_a^\mu)_{a \in \mathbb{A}}$ is a unitary representation of $\mathbb{A}$. Since $C_0(\widehat{\mathbb{A}})$ is dense in $L^2(\widehat{\mathbb{A}}, \mu)$, the latter space is separable. Therefore, also direct sums $\bigoplus_{i=1}^\infty V^{\mu_i}$ of such type representations will be unitary representations of $\mathbb{A}$ in separable Hilbert spaces.

Lemma 1 (Wiener Lemma) *If $F \subset L^2(\widehat{\mathbb{A}}, \mu)$ is a closed V_a^μ -invariant subspace for all $a \in \mathbb{A}$, then $F = 1_Y L^2(\widehat{\mathbb{A}}, \mathcal{B}(\widehat{\mathbb{A}}), \mu)$ for some Borel subset $Y \subset \widehat{\mathbb{A}}$.*

Notice however that since $\mathbb{A}$ is not compact (equivalently, $\widehat{\mathbb{A}}$ is not discrete), we can find μ continuous, and therefore V^μ has no irreducible (non-zero) subrepresentation. From now on, only unitary representations of $\mathbb{A}$ in separable Hilbert spaces will be considered, and we will show how to classify them.

A function $r : \mathbb{A} \rightarrow \mathbb{C}$ is called *positive definite* if

$$\sum_{n,m=0}^N r(a_n - a_m) z_n \overline{z_m} \geq 0 \quad (1)$$

for each $N > 0$, $(a_n) \subset \mathbb{A}$ and $(z_n) \subset \mathbb{C}$. The central result about positive definite functions is the following theorem (see, e.g., Rudin (1962)).

Theorem 1 (Bochner-Herglotz Theorem) *Let $r : \mathbb{A} \rightarrow \mathbb{C}$ be continuous. Then r is positive definite if and only if there exists (a unique) $\sigma \in M^+(\widehat{\mathbb{A}})$ such that*

$$r(a) = \int_{\widehat{\mathbb{A}}} \chi(a) d\sigma(\chi) \text{ for each } a \in \mathbb{A}.$$

If now $\mathcal{U} = (U_a)_{a \in \mathbb{A}}$ is a representation of $\mathbb{A}$ in H , then for each $x \in H$, the function $r(a) :=$

$\langle U_a x, x \rangle$ is continuous and satisfies (1), so it is positive definite. By the Bochner-Herglotz theorem, there exists a unique measure $\sigma_{u_x} = \sigma_x \in M^+(\widehat{\mathbb{A}})$ (called the *spectral measure* of x) such that

$$\widehat{\sigma}_x(a) := \int_{\widehat{\mathbb{A}}} i(a)(\chi) d\sigma_x(\chi) = \langle U_a x, x \rangle$$

for each $a \in \mathbb{A}$. Since the partial map $U_a x \mapsto i(a) \in L^2(\widehat{\mathbb{A}}, \sigma_x)$ is isometric and equivariant, there exists a unique extension of it to a unitary operator $W : \mathbb{A}(x) \rightarrow L^2(\widehat{\mathbb{A}}, \sigma_x)$ giving rise to an isomorphism of $\mathcal{U}|_{\mathbb{A}(x)}$ and V^{σ_x} . Then the existence of a spectral decomposition is proved by making use of separability and a choice of maximal cyclic spaces at every step of an induction procedure. Moreover, a spectral decomposition is unique in the following sense.

Theorem 2 (Spectral Theorem) *If $H = \oplus_{i=1}^{\infty} \mathbb{A}(x_i) = \oplus_{i=1}^{\infty} \mathbb{A}(x'_i)$ are two spectral decompositions of H , then $\sigma_{x_i} \equiv \sigma_{x'_i}$ for each $i \geq 1$.*

It follows that the representation $\mathcal{U}$ is entirely determined by the types (the sets of equivalent measures to a given one) of a decreasing sequence of measures, or, equivalently, $\mathcal{U}$ is determined by its maximal spectral type $\sigma_{\mathcal{U}}$ and its multiplicity function $M_{\mathcal{U}}$.

Notice that eigenvalues of $\mathcal{U}$ correspond to Dirac measures: $\chi \in \widehat{\mathbb{A}}$ is an eigenvalue (i.e., for some $\|x\| = 1$, $U_a(x) = \chi(a)x$ for each $a \in \mathbb{A}$) if and only if $\sigma_{u_x} = \delta_{\chi}$. Therefore, $\mathcal{U}$ has a discrete spectrum if and only if the maximal spectral type of $\mathcal{U}$ is a discrete measure.

We refer the reader to Glasner (2003), Katok and Thouvenot (2006), Lemańczyk (1996), Nadkarni (1998), and Parry (1981) for presentations of spectral theory needed in the theory of dynamical systems – such presentations are usually given for $\mathbb{A} = \mathbb{Z}$, but once we have the Bochner-Herglotz theorem and the Wiener lemma, their extensions to the general case are straightforward.

Koopman Representations

We will consider *measure-preserving* representations of $\mathbb{A}$. It means that we fix a probability standard Borel space $(X, \mathcal{B}, \mu)$, and by $\text{Aut}(X, \mathcal{B}, \mu)$, we denote the group of automorphisms of this space, that is, $T \in \text{Aut}(X, \mathcal{B}, \mu)$ if $T : X \rightarrow X$ is a bimeasurable (a.e.) bijection satisfying $\mu(A) = \mu(TA) = \mu(T^{-1}A)$ for each $A \in \mathcal{B}$. Consider then a representation of $\mathbb{A}$ in $\text{Aut}(X, \mathcal{B}, \mu)$, that is, a group homomorphism $a \mapsto T_a \in \text{Aut}(X, \mathcal{B}, \mu)$; we write $\mathcal{T} = (T_a)_{a \in \mathbb{A}}$. Moreover, we require that the associated Koopman representation $\mathcal{U}_{\mathcal{T}}$ is continuous. Unless explicitly stated, $\mathbb{A}$ -actions are assumed to be *free*, that is, we assume that for μ -a.e. $x \in X$ the map $a \mapsto T_a x$ is 1–1. In fact, since constant functions are obviously invariant for U_{T_a} , equivalently, the trivial character 1 is always an eigenvalue of $\mathcal{U}_{\mathcal{T}}$; the Koopman representation is considered only on the subspace $L_0^2(X, \mathcal{B}, \mu)$ of zero mean functions. We will restrict our attention only to *ergodic* dynamical systems (see Section “Glossary and Notation”). It is easy to see that $\mathcal{T}$ is ergodic if and only if whenever $A \in \mathcal{B}$ and $A = T_a(A)$ (μ -a.e.) for all $a \in \mathbb{A}$, then $\mu(A)$ equals 0 or 1. In case of ergodic Koopman representations, all eigenvalues are simple. In particular, (ergodic) Koopman representations with discrete spectra have simple spectra. The reader is referred to monographs mentioned above as well as to Cornfeld et al. (1982), Petersen (1983), Rudolph (1990), Sinai (1994), and Walters (1982) for basic facts on the theory of dynamical systems. See also survey articles (Lemańczyk 2009; Danilenko 2013).

The passage $\mathcal{T} \rightarrow \mathcal{U}_{\mathcal{T}}$ can be seen as functorial (contravariant). In particular, a measure-theoretic isomorphism of $\mathbb{A}$ -systems $\mathcal{T}$ and $\mathcal{T}'$ implies spectral isomorphism of the corresponding Koopman representations; hence, spectral properties are measure-theoretic invariants. Since unitary representations are completely classified, one of the main questions in the spectral theory of dynamical systems is to decide which pairs $([\sigma], M)$ can be realized by Koopman representations. The most spectacular is the Banach problem concerning a realization, for $\mathbb{A} = \mathbb{Z}$, of

$([\lambda_{\mathbb{T}}], M \equiv 1)$; see section “[Remarks on the Banach Problem](#).” Another important problem is to give a complete spectral classification in some classes of dynamical systems (classically, it was done in the theory of Kolmogorov and Gaussian dynamical systems). We will also see how spectral properties of dynamical systems can determine their statistical (ergodic) properties; a culmination given by the fact that a spectral isomorphism may imply measure-theoretic similitude (discrete spectrum case, Gaussian-Kronecker case). An old conjecture is that a dynamical system $\mathcal{T}$ either is spectrally determined or there are uncountably many pairwise non-isomorphic systems spectrally isomorphic to $\mathcal{T}$.

We could also consider Koopman representations in L^p for $1 \leq p \neq 2$. However, whenever $W: L^p(X, \mathcal{B}, \mu) \rightarrow L^p(Y, \mathcal{C}, \nu)$ is a surjective isometry and $W \circ U_{T_a} = U_{S_a} \circ W$ for each $a \in \mathbb{A}$, then by the Lamperti theorem (e.g., Royden 1968), the isometry W has to come from a non-singular pointwise map $R: Y \rightarrow X$, and, by ergodicity, R “preserves” the measure ν and hence establishes a measure-theoretic isomorphism (Kachurovskii 1990) (see also Lemańczyk (1996)). Thus, spectral classification of such Koopman representations goes back to the measure-theoretic classification of dynamical systems, so it looks hardly interesting. This does not mean that there are no interesting dynamical questions for $p \neq 2$. Let us mention still open Thouvenot’s question (formulated in the 1980s) for $\mathbb{Z}$ -actions: *For every ergodic Tacting on $(X, \mathcal{B}, \mu)$, does there exist $f \in L^1(X, \mathcal{B}, \mu)$ such that the closed linear span of $f \circ T^n$, $n \in \mathbb{Z}$, equals $L^1(X, \mathcal{B}, \mu)$?*

Iwanik (1991, 1992) proved that if T is a system with positive entropy, then its L^p -multiplicity is ∞ for all $p > 1$. Moreover, Iwanik and de Sam Lazaro (1991) proved that for Gaussian systems (they will be considered in section “[Spectral Theory of Dynamical Systems of Probabilistic Origin](#)”), the L^p -multiplicities are the same for all $p > 1$ (see also Lemańczyk and de Sam Lazaro (1997)).

Markov Operators, Joinings and Koopman Representations, Disjointness and Spectral Disjointness, and Entropy

We would like to emphasize that spectral theory is closely related to the theory joinings (see de la Rue’s article (de la Rue 2009) for needed definitions). The elements ρ of the set $J(\mathcal{S}, \mathcal{T})$ of joinings of two $\mathbb{A}$ -actions $\mathcal{S}$ and $\mathcal{T}$ are in a 1–1 correspondence with Markov operators $J = J_\rho$ between the L^2 -spaces equivariant with the corresponding Koopman representations (see section “[Glossary and Notation](#)” and de la Rue (2009)). The set of ergodic self-joinings of an ergodic $\mathbb{A}$ -action $\mathcal{T}$ is denoted by $J_2^e(\mathcal{T})$.

Each Koopman representation $\mathcal{U}_{\mathcal{T}}$ consists of Markov operators (indeed, U_{T_a} is clearly a Markov operator). In fact, if $U \in U(L^2(X, \mathcal{B}, \mu))$ is Markov, then it is of the form U_R , where $R \in \text{Aut}(X, \mathcal{B}, \mu)$; (Lemańczyk and Parreau 2012). This allows us to see Koopman representations as unitary Markov representations, but since a spectral isomorphism does not “preserve” the set of Markov operators, spectrally isomorphic systems can have drastically different sets of self-joinings.

We will touch here only some aspects of interactions (clearly, far from completeness) between the spectral theory and the theory of joinings.

In order to see however an example of such interactions, let us recall that the simplicity of eigenvalues for ergodic systems yields a short “joining” proof of the classical isomorphism theorem of Halmos-von Neumann in the discrete spectrum case: *Assume that $\mathcal{S} = (S_a)_{a \in \mathbb{A}}$ and $\mathcal{T} = (T_a)_{a \in \mathbb{A}}$ are ergodic $\mathbb{A}$ -actions on $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$, respectively. Assume that both Koopman representations have purely discrete spectrum and that their group of eigenvalues are the same. Then $\mathcal{S}$ and $\mathcal{T}$ are measure-theoretically isomorphic.* Indeed, each ergodic joining of $\mathcal{T}$ and $\mathcal{S}$ is the graph of an isomorphism of these two systems (see Lemańczyk (1996); see also Goodson’s proof in Goodson (1999)). Another example of such interactions appears in the study Rokhlin’s multiple mixing problem and its relation with the *pairwise independence*

property (PID) for joinings of higher order. We will not deal with this subject here, referring the reader to de la Rue (2009) (see however section “Lifting Mixing Properties”).

Following Furstenberg (1967), two $\mathbb{A}$ -actions $\mathcal{S}$ and $\mathcal{T}$ are called *disjoint* provided the product measure is the only element in $J(\mathcal{S}, \mathcal{T})$ (if they are disjoint, one of these actions has to be ergodic). It was already noticed in Hahn and Parry (1968) that spectrally disjoint systems are disjoint in the Furstenberg sense; indeed, $\text{Im} \left(J_\rho|_{L_0^2} \right) = \{0\}$ since $\sigma_{\mathcal{T}, J_\rho f} \ll \sigma_{\mathcal{S}, f}$.

Notice that whenever we take $\rho \in J_2^e(\mathcal{T})$, we obtain a new ergodic $\mathbb{A}$ -action $(T_a \times T_a)_{a \in \mathbb{A}}$ defined on the probability space $(X \times X, \rho)$. One can now ask how close spectrally to $\mathcal{T}$ is this new action? It turns out that except of the obvious fact that the marginal σ -algebras are factors, $(\mathcal{T} \times \mathcal{T}, \rho)$ can have other factors spectrally disjoint from $\mathcal{T}$: the most striking phenomenon here is a result of Smorodinsky and Thouvenot (1979) (see also Danilenko and Park (2002)) saying that each zero entropy system is a factor of an ergodic self-joining system of a fixed Bernoulli system (Bernoulli systems themselves have countable Haar spectrum). The situation changes if $\rho = \mu \otimes \mu$. In this case, for $f, g \in L^2(X, \mathcal{B}, \mu)$, the spectral measure of $f \otimes g$ is equal to $\sigma_{\mathcal{T}, f} * \sigma_{\mathcal{T}, g}$. A consequence of this observation is that an ergodic action $\mathcal{T} = (T_a)_{a \in \mathbb{A}}$ is weakly mixing (see section “Glossary and Notation”) if and only if the product measure $\mu \otimes \mu$ is an ergodic self-joining of $\mathcal{T}$.

The entropy which is a basic measure-theoretic invariant does not appear when we deal with spectral properties. We will not give here any formal definition of entropy for amenable group actions referring the reader to Ornstein and Weiss (1987). Assume that $\mathbb{A}$ is countable and discrete. We always assume that $\mathbb{A}$ is Abelian; hence, it is amenable. For each dynamical system $\mathcal{T} = (T_a)_{a \in \mathbb{A}}$ acting on $(X, \mathcal{B}, \mu)$, we can find a largest invariant sub- σ field $\mathcal{P} \subset \mathcal{B}$, called the *Pinsker σ -algebra*, such that the entropy of the corresponding quotient system is zero. Generalizing the classical Rokhlin-Sinai theorem (see also Kamiński (1981) for $\mathbb{Z}^d$ -actions), Thouvenot (unpublished) and independently Dooley and

Golodets (2002) proved this theorem for groups even more general than those considered here: *The spectrum of $\mathcal{U}_{\mathcal{T}}$ on $L^2(X, \mathcal{B}, \mu) \ominus L^2(\mathcal{P})$ is Haar with uniform infinite multiplicity*. This general result is quite intricate and based on methods introduced to entropy theory by Rudolph and Weiss (2000) with a very surprising use of Dye’s theorem on orbital equivalence of all ergodic systems. For $\mathbb{A}$ which is not countable, the same result was proved in Arni (2005) in case of unimodular amenable groups which are not increasing union of compact subgroups. It follows that spectral theory of dynamical systems essentially reduces to the zero entropy case.

Maximal Spectral Type of a Koopman Representation: Alexeyev’s Theorem

Only few general properties of maximal spectral types of Koopman representations are known. The fact that a Koopman representation preserves the space of real functions implies that its maximal spectral type is the type of a symmetric (invariant under the map $\chi \mapsto \bar{\chi}$) measure.

Recall that the *Gelfand spectrum* $\sigma(\mathcal{U})$ of a representation $\mathcal{U} = (U_a)_{a \in \mathbb{A}}$ is defined as the set of *approximative eigenvalues* of $\mathcal{U}$, that is, $\sigma(\mathcal{U}) \ni \chi \in \hat{\mathbb{A}}$ if for a sequence (x_n) bounded and bounded away from zero, $\|U_a x_n - \chi(a)x_n\| \rightarrow 0$ for each $a \in \mathbb{A}$. The spectrum is a closed subset in the topology of pointwise convergence, hence in the compact-open topology of $\hat{\mathbb{A}}$. In case of $\mathbb{A} = \mathbb{Z}$, the above set $\sigma(\mathcal{U})$ is equal to $\{z \in \mathbb{C} : U - z \cdot \text{Id} \text{ is not invertible}\}$.

Assume now that $\mathbb{A}$ is countable and discrete (and Abelian). Then there exists a Følner sequence $(B_n)_{n \geq 1}$ whose elements tile $\mathbb{A}$ (Ornstein and Weiss 1987). Take a free and ergodic action $\mathcal{T} = (T_a)_{a \in \mathbb{A}}$ on $(X, \mathcal{B}, \mu)$. By Ornstein and Weiss (1987) for each $\varepsilon > 0$, we can find a set $Y_n \in \mathcal{B}$ such that the sets $T_b Y_n$ are pairwise disjoint for $b \in B_n$ and $\mu(\cup_{b \in B_n} T_b Y_n) > 1 - \varepsilon$. For each $\chi \in \hat{\mathbb{A}}$, by considering functions of the form $f_n = \sum_{b \in B_n} \chi(b) 1_{T_b Y_n}$, we obtain that $\chi \in \sigma(\mathcal{U}_{\mathcal{T}})$. It follows that the topological support of the maximal spectral type of the Koopman representation of a free and ergodic

action is full (Katok and Thouvenot 2006; Lemańczyk 1996; Nadkarni 1998). The theory of Gaussian systems shows in particular that there are symmetric measures on the circle whose topological support is the whole circle but which cannot be maximal spectral types of Koopman representations.

An open well-known question remains of whether an absolutely continuous measure ρ is the maximal spectral type of a Koopman representation if and only if ρ is equivalent to a Haar measure of $\hat{\mathbb{A}}$ (this is unknown for $\mathbb{A} = \mathbb{Z}$).

Another interesting question was raised by A. Katok (see Katok and Lemańczyk (2009)): *Is it true that the topological supports of all measures in a spectral sequence of a Koopman representation are full?* If the answer to this question is positive, then, for example, the essential supremum of $M_{\mathcal{U}_T}$ is the same on all balls of $\hat{\mathbb{A}}$.

Notice that the fact that all spectral measures in a spectral sequence are symmetric means that $\mathcal{U}_T$ is isomorphic to $\mathcal{U}_{T^{-1}}$. A. del Junco (1981) showed that generically for $\mathbb{Z}$ -actions, T and its inverse are not measure-theoretically isomorphic (in fact, he proved disjointness).

Let $\mathcal{T}$ be an $\mathbb{A}$ -action on $(X, \mathcal{B}, \mu)$. One can ask whether a “good” function can realize the maximal spectral type of $\mathcal{U}_T$. In particular, can we find a function $f \in L^\infty(X, \mathcal{B}, \mu)$ that realizes the maximal spectral type? The answer is given in the following general theorem (see Lemańczyk and Wasieczko (2006)).

Theorem 3 (Alexeyev’s Theorem) *Assume that $\mathcal{U} = (U_a)_{a \in \mathbb{A}}$ is a unitary representation of $\mathbb{A}$ in a separable Hilbert space H . Assume that $F \subset H$ is a dense linear subspace. Assume moreover that with some F -norm $|\cdot|$ – stronger than the norm $\|\cdot\|$ given by the scalar product – F becomes a Fréchet space. Then, for each spectral measure $\sigma \ll \sigma_{\mathcal{U}}$, there exists $y \in F$ such that $\sigma_y \gg \sigma$. In particular, there exists $y \in F$ that realizes the maximal spectral type.*

By taking $H = L^2(X, \mathcal{B}, \mu)$ and $F = L^\infty(X, \mathcal{B}, \mu)$, we obtain the positive answer to the original question. Alexeyev (1958) proved the

above theorem for $\mathbb{A} = \mathbb{Z}$ using analytic functions. Refining Alexeyev’s original proof, Frączek (1997) showed the existence of a sufficiently regular function realizing the maximal spectral type depending only on the “regularity” of the underlying probability space, e.g., when X is a compact metric space (compact smooth manifold), then one can find a continuous (smooth) function realizing the maximal spectral type.

By the theory of systems of probabilistic origin (see section “Spectral Theory of Dynamical Systems of Probabilistic Origin”), in case of simplicity of the spectrum, one can easily point out spectral measures whose types are not realized by (essentially) bounded functions. However, it is still an open question whether for each Koopman representation $\mathcal{U}_T$ there exists a sequence $(f_i)_{i \geq 1} \subset L^\infty(X, \mathcal{B}, \mu)$ such that the sequence $(\sigma_{f_i})_{i \geq 1}$ is a spectral sequence for $\mathcal{U}_T$. For $\mathbb{A} = \mathbb{Z}$, it is unknown whether the maximal spectral type of a Koopman representation can be realized by a characteristic function.

The group $\text{Aut}(X, \mathcal{B}, \mu)$ considered with the weak operator topology is closed in $U(L^2(X, \mathcal{B}, \mu))$, hence becoming a Polish group (If we choose $\{A_i : i \geq 1\}$ a dense subset in $\mathcal{B}$ (considered modulo null sets), then the weak operator topology is metrizable with the metric $d(T_1, T_2) := \sum_{i \geq 1} \frac{1}{2^i} (\mu(T_1(A_i)\Delta T_2(A_i)) + \mu(T_1^{-1}(A_i)\Delta T_2^{-1}(A_i)))$). One can then ask what are “typical” (largeness is understood as a residual subset) properties of an automorphism of $(X, \mathcal{B}, \mu)$. It is classical (Halmos) that typically an automorphism is weakly mixing and rigid and has simple spectrum. Some other typical properties will be discussed later. While Halmos already noticed that in the weak operator topology mixing automorphisms form a meager set, in Tikhonov (2012), S. Tikhonov considers a special (Polish) topology on the set of mixing automorphisms. In fact, this topology was introduced by Alpern (1985) in 1985, and Tikhonov disproves a conjecture by Alpern by showing that a generic mixing transformation has simple singular spectrum and is mixing of arbitrary order; moreover, all its powers are disjoint. In Tikhonov (2013), the topology is extended to mixing actions of infinite

countable groups H ; it is given by the metric d_m , where for two H -actions $\mathcal{T}_i$ and $H \ni h \mapsto |h| \in \mathbb{N}$ so that $\sum_{h \in H} 1/2^{|h|} < +\infty$, we have

$$d_m(\mathcal{T}_1, \mathcal{T}_2) := \sum_{h \in H} \frac{1}{2^{|h|}} d(\mathcal{T}_{1,h}, \mathcal{T}_{2,h}) \\ + \sup_{h \in H_{i,j} \geq 1} \sum \frac{1}{2^{i+j}} \\ | \mu((\mathcal{T}_{1,h} A_i \cap A_j) - \mu(\mathcal{T}_{2,h} A_i \cap A_j)) |.$$

Bashtanov (2013, 2016) proved that the conjugacy classes (of mixing automorphisms) are dense in this topology. Hence, properties like to have trivial centralizer and no (non-trivial) factors are typical in this topology.

Spectral Theory of Weighted Operators

We now pass to the problem of possible essential values for the multiplicity function of a Koopman representation. However, one of the known techniques is a use of cocycles, so before we tackle the multiplicity problem, we will go through some results concerning spectral theory of compact group extension automorphisms which in turn entail a study of weighted operators (see section “Glossary and Notation”).

Assume that T is an ergodic automorphism of a standard Borel probability space $(X, \mathcal{B}, \mu)$. Let $\xi : X \rightarrow \mathbb{T}$ be a measurable function and let $V = V_{\xi, T}$ be the corresponding weighted operator. To see a connection of weighted operators with Koopman representations of compact group extensions, consider a compact (metric) Abelian group G and a cocycle $\varphi : X \rightarrow G$. Then U_{T_φ} (see section “Glossary and Notation”) acts on $L^2(X \times G, \mu \otimes \lambda_G)$. But

$$L^2(X \times G, \mu \otimes \lambda_G) = \bigoplus_{\chi \in \widehat{G}} L_\chi, \\ \text{where } L_\chi = L^2(X, \mu) \otimes \chi$$

is a U_{T_φ} -invariant (clearly, closed) subspace. Moreover, the map $f \otimes \chi \mapsto f$ settles a unitary isomorphism of $U_{T_\varphi}|_{L_\chi}$ with the operator $V_{\chi \circ \varphi, T}$.

Therefore, the spectral analysis of such Koopman representations reduces to the spectral analysis of weighted operators $V_{\chi \circ \varphi, T}$ for all $\chi \in \widehat{G}$.

Maximal Spectral Type of Weighted Operators over Rotations

The spectral analysis of weighted operators $V_{\xi, T}$ is especially well developed in case of rotations, i.e., when we additionally assume that T is an ergodic rotation on a compact monothetic group $X : Tx = x + x_0$, where x_0 is a topologically cyclic element of X (and μ will stand for Haar measure λ_X of X). In this case, Helson’s analysis (Helson 1986) applies (see also Gromov (1991), Iwanik et al. (1993), Lemańczyk (1996), and Queff  lec (1988)), leading to the following conclusions:

- The maximal spectral type $\sigma_{V_{\xi, T}}$ is either discrete or continuous.
- When $\sigma_{V_{\xi, T}}$ is continuous, it is either singular or Lebesgue.
- The spectral multiplicity of $V_{\xi, T}$ is uniform.

We now pass to a description of some results in case when $Tx = x + \alpha$ is an irrational rotation on the additive circle $X = [0, 1)$. It was already noticed in the original paper by Anzai (1951) that when $\xi : X \rightarrow \mathbb{T}$ is an affine cocycle ($\xi(x) = \exp(nx + c)$, $0 \neq n \in \mathbb{Z}$), then $V_{\xi, T}$ has a Lebesgue spectrum. It was then considered by several authors (originated by Kushnirenko (1974); see also Choe (1990) and Cornfeld et al. (1982)) to which extent this property is stable when we perturb our cocycle. Since the topological degree of affine cocycles is different from zero, when perturbing them, we consider smooth perturbations by cocycles of degree zero.

Theorem 4 Iwanik et al. (1993) *Assume that $Tx = x + \alpha$ is an irrational rotation. If $\xi : [0, 1) \rightarrow \mathbb{T}$ is of non-zero degree, absolutely continuous, with the derivative of bounded variation then $V_{\xi, T}$ has a Lebesgue spectrum.*

In the same paper, it is noticed that if we drop the assumption on the derivative, then the maximal spectral type of $V_{\xi, T}$ is a Rajchman measure

(i.e., its Fourier transform vanishes at infinity). It is still an open question whether one can find ξ absolutely continuous with non-zero degree and such that $V_{\xi,T}$ has singular spectrum. “Below” absolute continuity, topological properties of the cocycle seem to stop playing any role – in Iwanik et al. (1993), a continuous, degree 1 cocycle ξ of bounded variation is constructed such that $\xi(x) = \eta(x)/\eta(Tx)$ for a measurable $\eta: [0, 1) \rightarrow \mathbb{T}$ (i.e., ξ is a *coboundary*), and therefore $V_{\xi,T}$ has purely discrete spectrum (it is isomorphic to U_T). For other results about Lebesgue spectrum for Anzai skew products, see also Choe (1990), Frączek (2000), and Iwanik (1997); in Frączek (2000), $\mathbb{Z}^d$ -actions of rotations and so-called winding numbers instead of topological degree are considered. For recent generalizations, see Cecchi and Tiedra de Aldecoa (2016) and Tiedra de Aldecoa (2015a).

When the cocycle is still smooth but its degree is zero, the situation drastically changes. Given an absolutely continuous function $f: [0, 1) \rightarrow \mathbb{R}$, M. Herman (1979), using the Denjoy-Koksma inequality (see, e.g., Kuipers and Niederreiter (1974)), showed that $f_0^{(q_n)} \rightarrow 0$ uniformly (here $f_0 = f - \int_0^1 f d\lambda_{[0,1]}$ and (q_n) stands for the sequence of denominators of α). It follows that $T_{e^{2\pi i f}}$ is rigid and hence has a singular spectrum. B. Fayad (2002a) shows that this result is no longer true if one-dimensional rotation is replaced by a multi-dimensional rotation (his counterexample is in the analytic class). See also Lemańczyk and Mauduit (1994) for the singularity of spectrum for functions f whose Fourier transform satisfies $o\left(\frac{1}{|n|}\right)$ condition or to Iwanik et al. (1999), where it is shown that sufficiently small variation implies singularity of the spectrum.

A natural class of weighted operators arises when we consider Koopman operators of rotations on nil-manifolds. We only look at the particular example of such a rotation on a quotient of the Heisenberg group $(\mathbb{R}^3, *)$ (a general spectral theory of nil-actions was mainly developed by W. Parry (1970) – these actions have countable Lebesgue spectrum in the orthocomplement of the subspace of eigenfunctions), that is, take the nil-

manifold $\mathbb{R}^3/\mathbb{Z}^3$ on which we define the nilrotation $S((x, y, z) * \mathbb{Z}^3) = (\alpha, \beta, 0) * (x, y, z) * \mathbb{Z}^3$, where α , β , and 1 are rationally independent. It can be shown that S is isomorphic to the skew product defined on $[0, 1)^2 \times \mathbb{T}$ by

$$\begin{aligned} T_\varphi : (x, y, z) &\mapsto (x + \alpha, y + \beta, z \cdot e^{2\pi i \varphi(x, y)}) \\ &= (x + \alpha, y + \beta, z + \alpha y) * \mathbb{Z}^3, \end{aligned}$$

where $\varphi(x, y) = \alpha y - \psi(x + \alpha, y + \beta) + \psi(x, y)$ with $\psi(x, y) = x[y]$. Since nil-cocycles can be considered as a certain analog of affine cocycles for one-dimensional rotations, it would be nice to explain to what kind of perturbations the Lebesgue spectrum property is stable.

Yet another interesting problem which is related to the spectral theory of extensions given by so-called Rokhlin cocycles (see section “Rokhlin Cocycles”) arises, when given $f: [0, 1) \rightarrow \mathbb{R}$, we want to describe spectrally the one-parameter set of weighted operators $W_c := V_{e^{2\pi i c f}, T}$; here T is a fixed irrational rotation by α . As proved by quite sophisticated arguments in Iwanik et al. (1999), if we take $f(x) = x$, then for all noninteger $c \in \mathbb{R}$, the spectrum of W_c is continuous and singular (see also Gromov (1991) and Medina (1994), where some special α ’s are considered). It has been open for some time if at all one can find $f: [0, 1) \rightarrow \mathbb{R}$ such that for each $c \neq 0$, the operator W_c has a Lebesgue spectrum. The positive answer is given in Wysokinska (2004): For example, if $f(x) = x^{-(2+\varepsilon)}$ ($\varepsilon > 0$) and α has bounded partial quotients, then W_c has a Lebesgue spectrum for all $c \neq 0$. All functions with such a property considered in Wysokinska (2004) are non-integrable. It would be interesting to find an integrable f with the above property.

We refer to Goodson (1999) and the references therein for further results especially for transformations of the form $(x, y) \mapsto (x + \alpha, 1_{[0, \beta)}(x) + y)$ on $[0, 1) \times \mathbb{Z}/2\mathbb{Z}$. Recall however that earlier Katok and Stepin (1967) used this kind of transformations to give a first counterexample to the Kolmogorov group property (see section “Glossary and Notation”) for the spectrum.

The Multiplicity Problem for Weighted Operators over Rotations

In case of perturbations of affine cocycles, this problem was already raised by Kushnirenko (1974). Some significant results were obtained by M. Guenais. Before we state her results, let us recall a useful criterion to find an upper bound for the multiplicity: *If there exist $c > 0$ and a sequence $(F_n)_{n \geq 1}$ of cyclic subspaces of H such that for each $y \in H$, $\|y\| = 1$ we have $\liminf_{n \rightarrow \infty} \|proj_{F_n} y\|^2 \geq c$, then $\text{esssup}(M_U) \leq 1/c$ which follows from a well-known lemma of Chacon (1970; Cornfeld et al. 1982; King 1988; Lemańczyk 1996). Using this and a technique which is close to the idea of local rank one (see Ferenczi (1984) and King (1988)), M. Guenais (1998) proved a series of results on multiplicity which we now list.*

Theorem 5 *Assume that $Tx = x + \alpha$ and let $\xi : [0, 1) \mapsto \mathbb{T}$ be a cocycle.*

- (i) *If $\xi(x) = e^{2\pi i c x}$, then $M_{V_{\xi, T}}$ is bounded by $|c| + 1$.*
- (ii) *If ξ is absolutely continuous and ξ is of topological degree zero, then $V_{\xi, T}$ has a simple spectrum.*
- (iii) *If ξ is of bounded variation, then $M_{V_{\xi, T}} \leq \max(2, 2\pi \text{Var}(\xi)/3)$.*

Remarks on the Banach Problem

We already mentioned in section “Introduction” the Banach problem in ergodic theory, which is simply the question whether there exists a Koopman representation for $\mathbb{A} = \mathbb{Z}$ with simple Lebesgue spectrum. In fact no example of a Koopman representation with Lebesgue spectrum of finite multiplicity is known. Helson and Parry (1978) asked for the validity of a still weaker version: *Can one construct T such that U_T has a Lebesgue component in its spectrum whose multiplicity is finite?* Quite surprisingly in Helson and Parry (1978), they give a general construction yielding for each ergodic T a cocycle $\varphi : X \rightarrow \mathbb{Z}/2\mathbb{Z}$ such that the unitary operator U_{T_φ} has a Lebesgue spectrum in the orthocomplement of functions depending only on the X -coordinate. Then Mathew and Nadkarni (1984) gave

examples of cocycles over so-called dyadic adding machine for which the multiplicity of the Lebesgue component was equal to 2. In Lemańczyk (1988), this was generalized to so-called Toeplitz $\mathbb{Z}/2\mathbb{Z}$ -extensions of adding machines: For each even number k , we can find a two-point extension of an adding machine so that the multiplicity of the Lebesgue component is k . Moreover, it was shown that Mathew and Nadkarni’s constructions from Mathew and Nadkarni (1984) in fact are close to systems arising from number theory (like the famous Rudin-Shapiro sequence, e.g., Queffélec (1988)), relating the result about multiplicity of the Lebesgue component to results by Kamae and Queffélec (1988). Independently of Lemańczyk (1988), Ageev (1988) showed that one can construct two-point extensions of the Chacon transformation realizing (by taking powers of the extension) each even number as the multiplicity of the Lebesgue component. Contrary to all previous examples, Ageev’s constructions are weakly mixing.

Still an open question remains whether for $\mathbb{A} = \mathbb{Z}$ one can find a Koopman representation with the Lebesgue component of multiplicity 1 (or even odd).

In Guenais (1999), M. Guenais studies the problem of Lebesgue spectrum in the classical case of Morse sequences (see Keane (1968) as well as Kwiatkowski (1981)), where the problem of spectral classification in this class is studied). All dynamical systems arising from Morse sequences have simple spectra (Kwiatkowski 1981). It follows that if a Lebesgue component appears in a Morse dynamical system, it has multiplicity 1. Guenais (1999) using a Riesz product technique relates the Lebesgue spectrum problem with the still open problem (a variation of the classical Littlewood problem) of whether a construction of so-called L^1 -ultraflat trigonometric polynomials with coefficients ± 1 is possible (in the very recent preprint (Balister et al. 2019), the Littlewood problem of existence of uniformly flat trigonometric polynomials has been solved, but it is unclear whether it yields the ultraflatness condition). However, it is proved in Guenais (1999) that such a construction can be carried

out on some compact Abelian groups and it leads, for an Abelian countable torsion group $\mathbb{A}$, to a construction of an ergodic action of $\mathbb{A}$ with simple spectrum and a Haar component in its spectrum.

In Prikhodko (2013), A. Prikhodko published a construction of a rank one flow (see section “Rank-1 and Related Systems”) having Lebesgue spectrum. As rank one implies simple spectrum, the result yields solution of Banach problem for $\mathbb{A} = \mathbb{R}$. To carry out the construction, Prikhodko proved the following L^1 -ultraflat version of the Littlewood conjecture: *For all $0 < a < b$ and $n \geq 1$, there are polynomials $P_n(t) = \sum_{j=0}^{n-1} e^{2\pi i w_j^{(n)} t}$ for some real numbers $w_j^{(n)}$, so that $\|P_n\|_{L^1([a,b])} / \sqrt{n} \rightarrow 1$ when $n \rightarrow \infty$.* It seems however that some of the arguments in the paper are written too briefly and no further clarifying presentation of methods/results/ideas from Prikhodko (2013) has appeared so far. It would also be extremely nice to explain the status of (El Abdalaoui 2015) by H. El Abdalaoui, first posted on arXiv in 2015, which states the solution of the original Banach problem (i.e., in the conservative infinite measure-preserving category).

Lifting Mixing Properties

We now give one more example of interactions between spectral theory and joinings (see section “Introduction”) that gives rise to a quick proof of the fact that r -fold mixing property of T ($r \geq 2$) lifts to a weakly mixing compact group extension T_φ (the original proof of this fact is due to D. Rudolph (1985)). Indeed, to prove r -fold mixing for a mixing(=2-mixing) transformation S (acting on $(Y, \mathcal{C}, \nu)$), one has to prove that each weak limit of off-diagonal self-joinings (given by powers of S , see de la Rue (2009)) of order r is simply the product measure $\nu^{\otimes r}$. We need also a Furstenberg’s lemma (Furstenberg 1981) about relative unique ergodicity (RUE) of compact group extensions T_φ : *If $\mu \otimes \lambda_G$ is an ergodic measure for T_φ then it is the only (ergodic) invariant measure for T_φ whose projection on the first coordinate is μ .* Now, the result about lifting r -fold mixing to compact group extensions follows directly from the fact that whenever T_φ is weakly mixing, $(\mu \otimes \lambda_G)^{\otimes r}$ is an

ergodic measure (this approach was shown to the second author by A. del Junco). In particular, if T is mixing and T_φ is weakly mixing, then for each $\chi \in \widehat{G} \setminus \{1\}$, the maximal spectral type of $V_{\chi \circ \varphi, T}$ is Rajchman.

See section “Rokhlin Cocycles” for a generalization of the lifting result to Rokhlin cocycle extensions.

The Multiplicity Function

In this section, only $\mathbb{A} = \mathbb{Z}$ is considered. For other groups, even for $\mathbb{R}$, much less is known. Clearly, given an automorphism T , by inducing its Koopman $\mathbb{Z}$ -representation, we obtain a one-parameter group $(V_t)_{t \in \mathbb{R}}$ of unitary operators, which has precisely the same properties as the original one, except that we added the eigenvalues $n \in \mathbb{R}$. Moreover, classically, the induced Koopman representation is given by the suspension of T , i.e., by the special flow T^f (see section “Glossary and Notation”), where $f = 1$, whence it is also Koopman but is never weakly mixing. See Danilenko and Lemańczyk (2013) and Danilenko and Solomko (2010), where some of the results below proved for $\mathbb{A} = \mathbb{Z}$ have been extended to (weakly mixing) flows. See also the case of so-called *product $\mathbb{Z}^d$ -actions* (Filipowicz 1997) and (Solomko 2012) for general countable Abelian group actions.

Contrary to the case of maximal spectral type, it is rather commonly believed that there are no restrictions for the set of essential values of Koopman representations. In fact, if we drop the assumption that we consider the finite measure-preserving case and let ourselves consider μ σ -finite and infinite, Danilenko and Ryzhikov (2010, 2011) proved that all subsets of $\{1, 2, \dots\} \cup \{\infty\}$ are Koopman realizable (in the weak mixing and mixing class, respectively).

Cocycle Approach

We will only concentrate on some results of the last four decades. In 1983, refining an earlier idea of Oseledets from 1960, E.A. Robinson (1983) proved that for each $n \geq 1$, there exists an ergodic

transformation whose maximal spectral multiplicity is n . Another important result was proved in Robinson Jr (1986) (see also Katok (2003a)), where it is shown that given a finite set $M \subset \mathbb{N}$ containing 1 and closed under the least common multiple, one can find (even a weakly mixing) T so that the set of essential values of the multiplicity function equals M . This result was then extended in Goodson et al. (1992) to infinite sets and finally in Kwiatkowski Jr and Lemańczyk (1995) (see also Ageev (2001)) to all subsets $M \subset \mathbb{N}$ containing 1. In fact, as we have already noticed in the previous section, the spectral theory for compact Abelian group extensions is reduced to a study of weighted operators and then to comparing maximal spectral types for such operators. This leads to sets of the form

$$M(G, \nu, H) = \{ \#(\{ \chi \circ \nu^i : i \in \mathbb{Z} \} \cap \text{anil}(H)) : \chi \in \text{anil}(H) \}$$

($H \subset G$ is a closed subgroup and ν is a continuous group automorphism of G). Then an algebraic lemma has been proved in Kwiatkowski Jr and Lemańczyk (1995) saying that each set M containing 1 is of the form $M(G, \nu, H)$ and the techniques to construct “good” cocycles and a passage to “natural factors” yielded the following: *For each $M \subset \{1, 2, \dots\} \cup \{\infty\}$ containing 1, there exists an ergodic automorphism such that the set of essential values for its Koopman representation equals M .* See also Robinson Jr (1988) for the case of non-Abelian group extensions.

A similar in spirit approach (that means, a passage to a family of factors) is present in the paper of Ageev (2008) in which he first applies Katok’s analysis (see Katok (2003a, b)) for spectral multiplicities of the Koopman representation associated with Cartesian products $T^{\times k}$ for a generic transformation T . In a natural way, this approach leads to study multiplicities of tensor products of unitary operators. Roughly, the multiplicity is computed as the number of atoms (counted modulo obvious symmetries) for conditional measures (see Katok (2003a)) of a product measure over its convolution. Ageev (2008)

proved that for a typical automorphism T , the set of the values of the multiplicity function for $U_{T^{\times k}}$ equals $\{k, k(k-1), \dots, k!\}$ and then he just passes to “natural” factors for the Cartesian products by taking sets invariant under a fixed subgroup of permutations of coordinates. In particular, he obtains all sets of the form $\{2, 3, \dots, n\}$ on L_0^2 . He also shows that such sets of multiplicities are realizable in the category of mixing transformations. See also Ryzhikov (2009) for a realization of sets of the form $\{k, l, kl\}$, $\{k, l, m, kl, km, lm, klm\}$, etc.

A further progress was done in 2009–2012, when first Katok and Lemańczyk (2009) proved that each finite subset $M \subset \{1, 2, \dots\} \cup \{\infty\}$ containing 2 can be realized as the set of essential values of an ergodic automorphism which was then, by overcoming some algebraic difficulties, extended by Danilenko (2010, 2012) (in the mixing category) to all subsets containing 2.

Multiplicity for Gaussian and Poissonian Automorphisms

We refer the reader to section “Spectral Theory of Dynamical Systems of Probabilistic Origin” for the definition and basic properties of Gaussian and Poissonian automorphisms. Recall that given a Poissonian automorphism, there is a Gaussian automorphism spectrally isomorphic to it (whether the converse holds is unknown). In Gaussian case, the classical Girsanov’s theorem from 1950 asserts that the maximal spectral multiplicity in this case is either one or infinity (with a possibility that ∞ is not an essential value); see Kułaga and Parreau (2012) for an elegant proof of this theorem. What is the family of subsets appearing as sets of essential values of the multiplicity functions of Koopman operators given by Gaussian (and also Poissonian) automorphisms was studied by Danilenko and Ryzhikov in 2011. They prove the following remarkable results:

- This family contains all multiplicative sub-semigroups of $\mathbb{N}$.
- This family contains other sets than multiplicative sub-semigroups of $\mathbb{N}$.

The latter shows that Proposition 6.4.4 (multiplicative nature of M_T in the Gaussian case) claimed in the book by Katok and Thouvenot (2006) and also in Robinson (1986) is not true.

In the unpublished preprint (Ryzhikov 2014), Ryzhikov shows that all subsets containing ∞ are Gaussian “realizable” (even in the mixing category).

Rokhlin’s Uniform Multiplicity Problem

The Rokhlin multiplicity problem (see the book by Anosov (2003)) was, given $n \geq 2$, to construct an ergodic transformation with uniform multiplicity n on L_0^2 . A solution for $n = 2$ was independently given by Ageev (1999) and Ryzhikov (1999) (see also Anosov (2003) and Goodson (1999)), and in fact it consists in showing that for some T (actually, any T with simple spectrum for which $\frac{1}{2}(Id + U_T)$ is in the weak operator closure of the powers of U_T will do), the multiplicity of $T \times T$ is uniform and equal to 2 (see also section “Future Directions”). In Tikhonov (2011) and Ryzhikov and Troitskaya (2016), the case $n = 2$ is solved in case of mixing automorphisms (flows).

In Ageev (2005), Ageev proposed a new approach which consists in considering actions of “slightly non-Abelian” groups and showing that for a “typical” action of such a group, a fixed “direction” automorphism has a uniform multiplicity. Shortly after publication of Ageev (2005), Danilenko (2006), following Ageev’s approach, considerably simplified the original proof. We will present Danilenko’s arguments.

Fix $n \geq 1$. Denote $\bar{e}_i = (0, \dots, 1, \dots, 0) \in \mathbb{Z}^n$, $i = 1, \dots, n$. We define an automorphism L of $\mathbb{Z}^n$ setting $L(\bar{e}_i) = \bar{e}_{i+1}$, $i = 1, \dots, n-1$ and $L(\bar{e}_n) = \bar{e}_1$. Using L we define a semi-direct product $G := \mathbb{Z}^n \rtimes \mathbb{Z}$ defining multiplication as $(u, k) \cdot (w, l) = (u + L^k w, k + l)$. Put $e_0 = (0, 1)$, $e_i = (\bar{e}_i, 0)$, $i = 1, \dots, n$ (and $Le_i = (L\bar{e}_i, 0)$). Moreover, denote $e_{n+1} = e_0^n = (0, n)$. Notice that $e_0 \cdot e_i \cdot e_0^{-1} = Le_i$ for $i = 1, \dots, n$ ($L(e_{n+1}) = e_{n+1}$).

Theorem 6 (Ageev, Danilenko) *For every unitary representation $\mathcal{U}$ of G in a separable Hilbert space H , for which $U_{e_1 - L^r e_1}$ has no non-trivial fixed points for $1 \leq r < n$, the essential values of the multiplicity function for $U_{e_{n+1}}$ are contained in the set of multiples of n . If, in addition, U_{e_0} has a simple spectrum, then $U_{e_{n+1}}$ has uniform multiplicity n .*

It is then a certain work to show that the assumption of the second part of the theorem is satisfied for a typical action of the group G . Using a special (C, F) -construction with all the cut-and-stack parameters explicit, Danilenko (2006) was also able to show that each set of the form $k \cdot M$, where $k \geq 1$ and M is an arbitrary subset of natural numbers containing 1, is realizable as the set of essential values of a Koopman representation.

Tikhonov (2011) proved the existence of a mixing automorphism of uniform multiplicity n on L_0^2 for all $n \geq 1$.

Rokhlin Cocycles

We consider now a certain class of extensions which should be viewed as a generalization of the concept of compact group extensions. We will focus on $\mathbb{Z}$ -actions only.

Assume that T is an ergodic automorphism of $(X, \mathcal{B}, \mu)$. Let G be an l.c.s.c. Abelian group. Assume that this group acts on $(Y, \mathcal{C}, \nu)$, that is, we have a G -action $S = (S_g)_{g \in G}$ on $(Y, \mathcal{C}, \nu)$. Let $\varphi : X \rightarrow G$ be a cocycle. We then define an automorphism $T_{\varphi, S}$ of the space $(X \times Y, \mathcal{B} \otimes \mathcal{C}, \mu \otimes \nu)$ by

$$T_{\varphi, S}(x, y) = (Tx, S_{\varphi(x)}(y)).$$

Such an extension is called a *Rokhlin cocycle extension* (the map $x \mapsto S_{\varphi(x)}$ is called a *Rokhlin cocycle*). Such an operation generalizes the case of compact group extensions; indeed, when G is compact, the action of G on itself by rotations preserves Haar measure. (It is quite surprising, that when only we admit G non-Abelian, then, as shown in Danilenko and Lemańczyk (2005),

each ergodic extension of T has a form of a Rokhlin cocycle extension.) Ergodic and spectral properties of such extensions are examined in several papers: Glasner (1994), Glasner and Weiss (1989), Lemańczyk and Lesigne (2001), Lemańczyk et al. (2003), Lemańczyk and Parreau (2003, 2012), Robinson Jr (1992), and Rudolph (1986). Since in these papers rather joining aspects are studied (among other things in Lemańczyk and Lesigne (2001), Furstenberg's RUE lemma is generalized to this new context), we will mention here only few results, mainly spectral, following Lemańczyk and Lesigne (2001) and Lemańczyk and Parreau (2012). We will constantly assume that G is non-compact. As $\varphi : X \rightarrow G$ is then a cocycle with values in a non-compact group, the theory of such cocycles is much more complicated (see, e.g., Schmidt (1977)), and in fact the theory of Rokhlin cocycle extensions leads to interesting interactions between classical ergodic theory, the theory of cocycles, and the theory of non-singular actions arising from cocycles taking values in non-compact groups – especially, the Mackey action associated with φ plays a crucial role here (see the problem of invariant measures for $T_{\varphi,S}$ in Lemańczyk and Parreau (2003) and Danilenko and Lemańczyk (2005); see also monographs by Aaronson (1979, 1997) Katok (2001, 2003a), and Schmidt (1977)). Especially, two Borel subgroups of $\widehat{G}$ are important here:

$$\Sigma_{\varphi} = \left\{ \chi \in \widehat{G} : \chi \circ \varphi = c \cdot \xi / \xi \circ T \text{ for a measurable } \xi : X \rightarrow \mathbb{T} \text{ and } c \in \mathbb{T} \right\}$$

and its subgroup Λ_{φ} given by $c = 1$. Λ_{φ} turns out to be the group of L^{∞} -eigenvalues of the Mackey action (of G) associated with the cocycle φ . This action is the quotient action of the natural action of G (by translations on the second coordinate) on the space of ergodic components of the skew product T_{φ} – the Mackey action is (in general) not measure-preserving; it is however non-singular. We refer the reader to Aaronson and Nadkarni (1987), Host et al. (1991), and Nadkarni (1998) for other properties of those subgroups.

Theorem 7 (Lemańczyk and Parreau (2003, 2012))

- (i) $\sigma_{T_{\varphi,S}}|_{L^2(X \times Y, \mu \otimes \nu) \ominus L^2(X, \mu)} = \int_{\widehat{G}} \sigma_{V_{\chi \circ \varphi, T}} d\sigma_S$.
- (ii) $T_{\varphi,S}$ is ergodic if and only if T is ergodic and $\sigma_S(\Lambda_{\varphi}) = 0$.
- (iii) $T_{\varphi,S}$ is weakly mixing if and only if T is weakly mixing and S has no eigenvalues in Σ_{φ} .
- (iv) If T is mixing, S is mildly mixing, and φ is recurrent and not cohomologous to a cocycle with values in a compact subgroup of G , then $T_{\varphi,S}$ remains mixing.
- (v) If T is r -fold mixing, φ is recurrent, and $T_{\varphi,S}$ is mildly mixing, then $T_{\varphi,S}$ is also r -fold mixing.
- (vi) If T and R are disjoint, the cocycle φ is ergodic, and S is mildly mixing, then $T_{\varphi,S}$ remains disjoint from R .

Let us recall (Furstenberg and Weiss 1978; Schmidt and Walters 1982) that an $\mathbb{A}$ -action $S = (S_a)_{a \in \mathbb{A}}$ is mildly mixing (see section “Glossary and Notation”) if and only if the $\mathbb{A}$ -action $(S_a \times \tau_a)_{a \in \mathbb{A}}$ remains ergodic for every properly ergodic non-singular $\mathbb{A}$ -action $\tau = (\tau_a)_{a \in \mathbb{A}}$.

Coming back to Smorodinsky-Thouvenot's result about factors of ergodic self-joinings of a Bernoulli automorphism, we would like to emphasize here that the disjointness result (vi) above was used in Lemańczyk and Parreau (2003) to give an example of an automorphism which is disjoint from all weakly mixing transformations but which has an ergodic self-joining whose associated automorphism has a non-trivial weakly mixing factor. In a sense, this is opposed to Smorodinsky-Thouvenot's result as here from self-joinings, we produced a “more complicated” system (namely, the weakly mixing factor) than the original system.

It would be interesting to develop the theory of spectral multiplicity for Rokhlin cocycle extensions as it was done in the case of compact group extensions. However, a difficulty is that in the compact group extension case, we deal with a countable direct sum of representations of the form $V_{\chi \circ \varphi, T}$ while in the non-compact case, we have to consider an integral of such representations.

Rank-1 and Related Systems

Although properties like mixing, weak (and mild) mixing, as well as ergodicity are clearly spectral properties, they have “good” measure-theoretic formulations (expressed by a certain behavior on sets). Simple spectrum property is another example of a spectral property, and it was a popular question in the 1980s whether simple spectrum property of a Koopman representation can be expressed in a more “measure-theoretic” way. We now recall rank-1 concept which can be seen as a notion close to Katok’s and Stepin’s theory of cyclic approximation (Katok and Stepin 1967) (see also Cornfeld et al. (1982)).

Assume that T is an automorphism of a standard probability Borel space $(X, \mathcal{B}, \mu)$. T is said to have the *rank one* property if there exists an increasing sequence of Rokhlin towers tending to the partition into points (a *Rokhlin tower* is a family $\{F_n, TF_n, \dots, T^{h_n-1}F_n\}$ of pairwise disjoint sets, while “tending to the partition into points” means that we can approximate every set in $\mathcal{B}$ by unions of levels of towers in the sequence). Hence, basically, rank one automorphism is given by two sequences of parameters: $r_n, n \geq 1$, which is the number of subcolumns on which we divide the n th tower given by F_n , and $S_{n,j}, n \geq 1, j = 0, 1, \dots, r_n - 1$, the sequence of spacers put over consecutive subcolumns. A “typical” automorphism of a standard probability Borel space has the rank one property. The “typicality” of rank one is still true in the Alpern-Tikhonov topology we mentioned in section “Maximal Spectral Type of a Koopman Representation: Alexeyev’s Theorem” by Bashtanov (2013).

Baxter (1971) showed that the maximal spectral type of such a T is realized by a characteristic function. Since the cyclic space generated by the characteristic function of the base contains characteristic functions of all levels of the tower, by the definition of rank one, the increasing sequence of cyclic spaces tends to the whole L^2 -space; therefore, rank one property implies simplicity of the spectrum for the Koopman representation. It was a question for some time whether rank-1 is just a characterization of simplicity of the spectrum, disproved by del Junco (1977). We refer the

reader to Ferenczi (1997) as a good source for basic properties of rank-1 transformations.

Similar to the rank one property, one can define *finite rank* automorphisms (simply by requiring that an approximation is given by a sequence of a fixed number of towers) – see, e.g., Ornstein et al. (1982), or even a more general property, namely, the *local rank one* property, can be defined, just by requiring that the approximating sequence of single towers fills up a fixed fraction of the space (see Ferenczi (1984) and King (1988)). Local rank one property implies finite multiplicity (King 1988), and the maximal spectral multiplicity is always bounded by rank. Mentzen (1988) showed that for each $n \geq 1$, one can construct an automorphism with simple spectrum and having rank n , and later Kwiatkowski and Lacroix (1997) showed that for each pair (m, r) with $m \leq r$, one can construct a rank r automorphism whose maximal spectral multiplicity is m . In Lemańczyk and Sikorski (1987), there is an example of a simple spectrum automorphism which is not of local rank one. Ferenczi (1985) deals with the notion of funny rank one (approximating towers are Rokhlin towers with “holes”) – the concept that has been introduced by Thouvenot. Funny rank one also implies simplicity of the spectrum. An example is given in Ferenczi (1985) which is even not loosely Bernoulli (see section “Inducing and Spectral Theory”; we recall that local rank one property implies loose Bernoullicity (Ferenczi 1984)).

The notion of AT (see section “Glossary and Notation”) has been introduced by Connes and Woods (1985). They proved that AT property implies zero entropy. They also proved that funny rank one automorphisms are AT. In Dooley and Quas (2005), it is proved that the system induced by the classical Morse-Thue system is AT (it is an open question by S. Ferenczi whether this system has funny rank one property). A question by Dooley and Quas is whether AT implies funny rank one property. AT property implies “simplicity of the spectrum in L^1 ” which we already considered in section “Introduction” (a “generic” proof of this fact is due to J.-P. Thouvenot).

A persistent question was formulated in the 1980s whether rank one itself is a spectral

property. In Ferenczi and Lemańczyk (1991), the authors maintained that this is not the case, based on an unpublished preprint of the first named author of Ferenczi and Lemańczyk (1991) in which there was a construction of a Gaussian-Kronecker automorphism (see section “Spectral Theory of Dynamical Systems of Probabilistic Origin”) having rank-1 property. This latter construction turned out to be false. In fact, de la Rue (1998a) proved that no Gaussian automorphism can be of local rank one. Therefore, the question whether *rank one is a spectral property* remains one of the interesting open questions in that theory. Downarowicz and Kwiatkowski (2000) proved that rank-1 is a spectral property in the class of systems generated by generalized Morse sequences.

One of the most beautiful theorems about rank-1 automorphisms is the following result of J. King (1986) (for a different proof, see Ryzhikov (1992)).

Theorem 8 WCT *If T is of rank one, then for each element S of the centralizer $C(T)$ of T , there exists a sequence (n_k) such that $U_T^{n_k} \rightarrow U_S$ strongly.*

A conjecture of J. King is that in fact for rank-1 automorphisms, each indecomposable Markov operator $J = J_\rho (\rho \in J_2^e(T))$ is a weak limit of powers of U_T (see King (2001) and also Ryzhikov (1992)). To which extent the WCT remains true for actions of other groups is not clear. In Zeitz (1993), the WCT is proved in case of rank one flows; however, the main argument seems to be based on the fact that a rank one flow has a non-zero time automorphism T_{t_0} which is of rank one, which is not true. After the proof of the WCT by Ryzhikov in Ryzhikov (1992), there is a remark that the rank one flow version of the theorem can be proved by a word-for-word repetition of the arguments. He also proves that if the flow $(T_t)_{t \in \mathbb{R}}$ is mixing, then T_1 does not have finite rank. On the other hand, for $\mathbb{A} = \mathbb{Z}^2$, Downarowicz and Kwiatkowski (2002) gave a counterexample to the WCT. But see also Janvresse et al. (2012).

Even though it looks as if rank one construction is not complicated, mixing in this class is possible; historically the first mixing constructions were given by D. Ornstein (1970) in 1970, using probability type arguments for a choice of spacers. Once mixing was shown, the question arose whether absolutely continuous spectrum is also possible, as this would give automatically the positive answer to the Banach problem. However, Bourgain (Bourgain 1993), relating spectral measures of rank one automorphisms with some classical constructions of Riesz product measures, proved that a certain subclass of Ornstein’s class consists of automorphisms with singular spectrum (see also El Abdalaoui (2007) and El Abdalaoui et al. (2006)). Since in Ornstein’s class spacers are chosen in a certain “non-constructive” way, quite a lot of attention was devoted to the rank one automorphism defined by cutting a tower at the n -th step into $r_n = n$ subcolumns of equal “width” and placing i spacers over the i -th subcolumn. The mixing property conjectured by M. Smorodinsky was proved by Adams (1998) (in fact Adams proved a general result on mixing of a class of staircase transformations). Spectral properties of rank-1 transformations are also studied in Klemes and Reinhold (1997), where the authors proved that whenever $\sum_{n=1}^{\infty} r_n^{-2} = +\infty$, then the spectrum is automatically singular (see also the more recent in Creutz and Silva (2010)). H. El Abdalaoui (2007) gives a criterion for singularity of the spectrum of a rank one transformation; his proof uses a central limit theorem. It seems that still the question whether rank one implies singularity of the spectrum remains the most important question of this theory.

We have already seen in section “Spectral Theory of Weighted Operators” that for a special class of rank one systems, namely, those with discrete spectra (del Junco 1976), we have a nice theory for weighted operators. It would be extremely interesting to find a rank one automorphism with continuous spectrum for which a substitute of Helson’s analysis exists.

B. Fayad (2005) constructs a rank one differentiable flow, as a special flow over a two-dimensional rotation. In Fayad (2006), he gives

new constructions of smooth flows with singular spectra which are mixing (with a new criterion for a Rajchman measure to be singular). In Fayad (2001a), a certain smooth change of time for an irrational flows on the 3-torus is given, so that the corresponding flow is partially mixing and has the local rank one property.

Motivated by Sarason's conjecture on Möbius disjointness (see Kułaga-Przymus and Lemańczyk (2019)), a certain recent activity was to study spectral disjointness of powers for rank one automorphisms. Let σ be a probability measure on the additive circle $[0, 1)$. Given a real number $a > 0$, we denote by σ^a the image of σ under the map $x \mapsto ax \bmod 1$. If $r \geq 1$ is an integer, then by σ_r , we will denote the measure which is obtained first by taking the image of σ under the map $x \mapsto \frac{1}{r}x$, i.e., the measure $\sigma^{1/r}$, and then repeating this new measure periodically in intervals $[\frac{j}{r}, \frac{j+1}{r})$. The following holds (e.g., in the unpublished notes by H. El Abdalaoui, J. Kułaga-Przymus, M. Lemańczyk, and T. de la Rue.):

if $(r, s) = 1$ then $\sigma^r \perp \eta^s$ if and only if $\sigma_s \perp \eta_r$.
(2)

In Bourgain (2013), Bourgain used Riesz product technique to show that for the class of so-called rank one automorphisms with bounded parameters (both (r_n) and $(s_{n,j})$ are bounded and no spacer over the last column), we have $\sigma_r \perp \sigma_s$ for $r \neq s$ prime. In view of (2), it follows that different prime powers are spectrally disjoint. In El Abdalaoui et al. (2014), a much larger class of rank one automorphisms is considered. No boundedness assumption on (r_n) is made, but a certain bounded recurrence is required on the sequence of spacers. Spectral disjointness of different powers (for the continuous part of the maximal spectral type) is derived from the existence, in the weak closure of powers, of sufficiently many analytic functions of the Koopman operator U_T .

For a spectral disjointness of the continuous part of the maximal spectral type for powers of automorphisms like the substitutional system

given by the Thue-Morse sequence and related (rank two systems), see El Abdalaoui et al. (2016). Weak closure of powers for Chacon automorphism is described in Janvresse et al. (2015).

Spectral Theory of Dynamical Systems of Probabilistic Origin

Let us just recall that when $(Y_n)_{n=-\infty}^{\infty}$ is a stationary process, then its distribution μ on $\mathbb{R}^{\mathbb{Z}}$ is invariant under the shift S on $\mathbb{R}^{\mathbb{Z}}$: $S((x_n)_{n \in \mathbb{Z}}) = (y_n)_{n \in \mathbb{Z}}$, where $y_n = x_{n+1}$, $n \in \mathbb{Z}$. In this way, we obtain an automorphism S defined on $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}(\mathbb{R}^{\mathbb{Z}}), \mu)$. For each automorphism T , we can find $f: X \rightarrow \mathbb{R}$ measurable such that the smallest σ -algebra making the stationary process $(f \circ T^n)_{n \in \mathbb{Z}}$ measurable is equal to $\mathcal{B}$; therefore, for the purpose of this entry, by a system of probabilistic origin, we will mean (S, μ) obtained from a stationary infinitely divisible process (see, e.g., Maruyama (1970) and Sato (1999)). In particular, the theory of Gaussian dynamical systems is indeed a classical part of ergodic theory (e.g., Newton 1966; Newton and Parry 1966; Vershik 1962a, b). If $(X_n)_{n \in \mathbb{Z}}$ is a stationary real centered Gaussian process and σ denotes the *spectral measure of the process*, i.e., $\widehat{\sigma}(n) = E(X_n \cdot X_0)$, $n \in \mathbb{Z}$, then by $S = S_\sigma$, we denote the corresponding Gaussian system on the shift space (recall also that for each symmetric measure σ on $\mathbb{T}$, there is exactly one stationary real centered Gaussian process whose spectral measure is σ). Notice that if σ has an atom, then in the cyclic space generated by X_0 , there exists an eigenfunction Y for S_σ – if now S_σ were ergodic, $|Y|$ would be a constant function which is not possible by the nature of elements in $\mathbb{Z}(X_0)$. In what follows, we assume that σ is continuous.

It follows that U_{S_σ} restricted to $\mathbb{Z}(X_0)$ is spectrally the same as $V = V^\sigma$ acting on $L^2(\mathbb{T}, \sigma)$, and we obtain that $(U_{S_\sigma}, L^2(\mathbb{R}^{\mathbb{Z}}, \mu_\sigma))$ can be represented as the symmetric Fock space built over $H = L^2(\mathbb{T}, \sigma)$ and $U_{S_\sigma} = F(V)$ – see section “Glossary and Notation” ($H^{\odot n}$ is called the n -th chaos). In other words, the spectral theory of Gaussian dynamical systems is reduced to the spectral theory of special tensor products unitary

operators. Classical results (see Cornfeld et al. (1982)) which can be obtained from this point of view are, for example, the following:

- (A) Ergodicity implies weak mixing.
- (B) The multiplicity function is either 1 or is unbounded.
- (C) The maximal spectral type of U_{S_σ} is equal to $\exp(\sigma)$; hence, Gaussian systems enjoy the Kolmogorov group property.

However, we can also look at a Gaussian system in a different way, simply by noticing that the variables $e^{2\pi if}$ (f is a real variable), where $f \in \mathbb{Z}(X_0)$ generate $L^2(\mathbb{R}^\mathbb{Z}, \mu_\sigma)$. Now, calculating the spectral measure of $e^{2\pi if}$ is not difficult and we obtain easily (C). Moreover, integrals of type $\int e^{2\pi if_0} e^{2\pi if_1 \circ \sigma^n} e^{2\pi if_2 \circ \sigma^{n+m}} d\mu_\sigma$ can also be calculated; whence, in particular, we easily obtain Leonov's theorem on the multiple mixing property of Gaussian systems (Leonov 1960).

One of the most beautiful parts of the theory of Gaussian systems concerns ergodic properties of S_σ when σ is concentrated on a thin Borel set. Recall that a closed subset $K \subset T$ is said to be a *Kronecker set* if each $f \in C(K)$ is a uniform limit of characters (restricted to K). Each Kronecker set has no rational relations. Gaussian-Kronecker automorphisms are, by definition, those Gaussian systems for which the measure σ (always assumed to be continuous) is concentrated on $K \cup \overline{K}$, K a Kronecker set. The following theorem has been proved in Foiaş and Stratila (1968) (see also Cornfeld et al. (1982)).

Theorem 9 Foiaş-Stratila Theorem *If T is an ergodic automorphism and f is a real-valued element of L_0^2 such that the spectral measure σ_f is concentrated on $K \cup \overline{K}$, where K is a Kronecker set, then the process $(f \circ T^n)_{n \in \mathbb{Z}}$ is Gaussian.*

This theorem is indeed striking as it gives examples of weakly mixing automorphisms which are spectrally determined (like rotations). A relative version of the Foiaş-Stratila theorem

has been proved in Lemańczyk and Lesigne (2001).

The Foiaş-Stratila theorem implies that whenever a spectral measure σ is Kronecker, it has no realization of the form σ_f with f bounded. We will see however in section “Future Directions” that for some automorphisms T (having the SCS property), the maximal spectral type σ_T has the property that S_{σ_T} has a simple spectrum.

Gaussian-Kronecker automorphisms are examples of automorphisms with simple spectra. In fact, whenever σ is concentrated on a set without rational relations, then S_σ has a simple spectrum (see Cornfeld et al. (1982)). Examples of mixing automorphisms with simple spectra are known (Newton 1966); however, it is still unknown (Thouvenot's question) whether the Foiaş-Stratila property may hold in the mixing class. F. Parreau (2000) using independent Helson sets gave an example of mildly mixing Gaussian system with the Foiaş-Stratila property.

Joining theory of a class of Gaussian system, called GAG, is developed in Lemańczyk et al. (2000). A Gaussian automorphism S_σ with the Gaussian space $H \subset L_0^2(\mathbb{R}^\mathbb{Z}, \mu_\sigma)$ is called a GAG if for each ergodic self-joining $\rho \in J_2^e(S_\sigma)$ and arbitrary $f, g \in H$ the variable

$$(\mathbb{R}^\mathbb{Z} \times \mathbb{R}^\mathbb{Z}, \rho) \ni (x, y) \mapsto f(x) + g(y)$$

is Gaussian. For GAG systems one can describe the centralizer and factors, they turn out to be objects close to the probability structure of the system. One of the crucial observations in Lemańczyk et al. (2000) was that all Gaussian systems with simple spectrum are GAG.

It is conjectured (J.P. Thouvenot) that in the class of zero entropy Gaussian systems, the PID property holds true.

For the spectral theory of classical factors of a Gaussian system, see Lemańczyk and de Sam Lazaro (1997); also spectrally they share basic spectral properties of Gaussian systems. Recall that historically one of the classical factors, namely, the σ -algebra of sets invariant for the map

$$\begin{aligned}
 (\dots, x_{-1}, x_0, x_1, \dots) &\mapsto \\
 (\dots, -x_{-1}, -x_0, -x_1, \dots),
 \end{aligned}$$

was the first example with zero entropy and countable Lebesgue spectrum (indeed, we need a singular measure σ such that $\sigma * \sigma$ is equivalent to Lebesgue measure (Newton and Parry 1966)). For factors obtained as functions of a stationary process, see Iwanik et al. (1997).

T. de la Rue (1998a) proved that Gaussian systems are never of local rank-1; however, his argument does not apply to classical factors. We conjecture that Gaussian systems are disjoint from rank-1 automorphisms (or even from local rank-1 systems).

We now turn the attention to Poissonian systems (see Cornfeld et al. (1982) for more details). Assume that $(X, \mathcal{B}, \mu)$ is a standard Borel space, where μ is infinite. Without entering too much into details, the new configuration space $\tilde{X}$ is taken as the set of all countable subsets $\{x_i : i \geq 1\}$ of X . Once a set $A \in \mathcal{B}$ of finite measure is given, one can define a map $N_A : \tilde{X} \rightarrow \mathbb{N}(\cup\{\infty\})$ just counting the number of elements belonging to A . The measure-theoretic structure $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu})$ is given so that the maps N_A become random variables with Poisson distribution of parameter $\mu(A)$ and such that whenever $A_1, \dots, A_k \subset X$ are of finite measure and are pairwise disjoint, then the variables $N_{A_1}, \dots, N_{A_k}$ are independent.

Assume now that T is an automorphism of $(X, \mathcal{B}, \mu)$. It induces a natural automorphism on the space $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu})$ defined by $\tilde{T}(\{x_i : i \geq 1\}) = \{Tx_i : i \geq 1\}$. The automorphism $\tilde{T}$ is called the *Poisson suspension* of T (see Cornfeld et al. (1982)). Such a suspension is ergodic if and only if no set of positive and finite μ -measure is T -invariant. Moreover, the ergodicity of $\tilde{T}$ implies weak mixing. In fact the spectral structure of $U_{\tilde{T}}$ is very similar to the Gaussian one: Namely, the first chaos equals $L^2(X, \mathcal{B}, \mu)$ (we emphasize that this is about the whole L^2 and not only L^2_0) on which $U_{\tilde{T}}$ acts as U_T and the $L^2(\tilde{X}, \tilde{\mu})$ together with the action of $U_{\tilde{T}}$ has the structure of the symmetric Fock space $\tilde{F}(L^2(X, \mathcal{B}, \mu))$ (see section “Glossary and Notation”).

We refer to Cambanis et al. (1995), Janicki and Weron (1994), and Rosiński and Żak (1996, 1997) for ergodic properties of systems given by symmetric α -stable stationary processes, or more generally infinitely divisible processes. Again, they share spectral properties similar to the Gaussian case: Ergodicity implies weak mixing, while mixing implies mixing of all orders.

In Roy (2007), E. Roy clarifies the dynamical “status” of such systems. He uses Poisson suspension automorphisms and the Maruyama representation of an infinitely divisible process mixed with basic properties of automorphisms preserving infinite measure (see Aaronson (1997)) to prove that as a dynamical system, a stationary infinitely divisible process (without the Gaussian part) is a factor of the Poisson suspension over the Lévy measure of this process. In Roy (2005), a theory of ID-joinings is developed (which should be viewed as an analog of the GAG theory in the Gaussian class). Parreau and Roy (2015) study Poisson suspensions without non-trivial factors.

Many natural problems still remain open here, for example (assuming always zero entropy of the dynamical system under consideration): Are Poisson suspensions disjoint from Gaussian systems? In Janvresse et al. (2017), there are examples of Poissonian systems which are disjoint from all Gaussian systems. What is the spectral structure for dynamical systems generated by symmetric α -stable processes? Are such systems disjoint whenever $\alpha_1 \neq \alpha_2$? Are Poissonian systems disjoint from local rank one automorphisms (cf. de la Rue 1998a)? In del Junco and Lemańczyk (1999), it is proved that Gaussian systems are disjoint from so-called simple systems (see Veech (1982a), del Junco and Rudolph (1987), and de la Rue (2009)); we will come back to an extension of this result in section “Future Directions.” It seems that flows of probabilistic origin satisfy the Kolmogorov group property for the spectrum. One can therefore ask how different are systems satisfying the Kolmogorov group property from systems for which the convolutions of the maximal spectral type are pairwise disjoint (see also section “Future Directions” and the SCS property).

We also mention here another problem which will be taken up in section “[Special Flows, Flows on Surfaces, and Interval Exchange Transformations](#)” – *Is it true that flows of probabilistic origin are disjoint from smooth flows on surfaces?*

Yet one more (joining) property seems to be characteristic in the class of systems of probabilistic origin, namely, they satisfy so-called ELF property (see Derriennic et al. (2008) and de la Rue’s article (de la Rue (2009))). Vershik asked whether the ELF property is spectral – however, the example of a cocycle from Wysokinska (2004) together with Theorem 7 (i) yields a certain Rokhlin extension of a rotation which is ELF and has countable Lebesgue spectrum in the orthocomplement of the eigenfunctions (see Wysokińska (2007)); on the other hand, any affine extension of that rotation is spectrally the same, while it cannot have the ELF property.

Prikhodko and Thouvenot (private communication) have constructed weakly mixing and non-mixing rank one automorphisms which enjoy the ELF property.

Inducing and Spectral Theory

Assume that T is an ergodic automorphism of a standard probability Borel space $(X, \mathcal{B}, \mu)$. Can “all” dynamics be obtained by inducing (see section “[Glossary and Notation](#)”) from one fixed automorphism was a natural question from the very beginning of ergodic theory. Because of Abramov’s formula for entropy $h(T_A) = h(T)/\mu(A)$, it is clear that positive entropy transformations cannot be obtained from inducing on a zero entropy automorphism. However, here we are interested in spectral questions, and thus we ask how many spectral types we obtain when we induce. It is proved in Friedman and Ornstein (1973) that the family of $A \in \mathcal{B}$ for which T_A is mixing is dense for the (pseudo) metric $d(A_1, A_2) = \mu(A_1 \Delta A_2)$. De la Rue (1998b) proves the following result: *For each ergodic transformation T of a standard probability space $(X, \mathcal{B}, \mu)$, the set of $A \in \mathcal{B}$ for which the maximal spectral type of U_{T_A} is Lebesgue is dense in $\mathcal{B}$.* The multiplicity

function is not determined in that paper. Recall (without giving a formal definition, see Ornstein et al. (1982)) that a zero entropy automorphism is *loosely Bernoulli* (LB for short) if and only if it can be induced from an irrational rotation (see also Feldman (1976) and Katok (1977)). The LB theory shows that not all dynamical systems can be obtained by inducing from an ergodic rotation. However, an open question remained whether LB systems exhaust spectrally all Koopman representations. An interesting question of M. Ratner (Ratner 1978) is whether from every ergodic automorphism T , one can induce an automorphism which has countable Lebesgue spectrum (Ratner in Ratner (1978) shows that this can be done if T is an irrational rotation).

In a deep paper (de la Rue 1996), de la Rue studies LB property in the class of Gaussian-Kronecker automorphisms, in particular he constructs S which is not LB. Suppose now that T is LB and for some $A \in \mathcal{B}$, U_{T_A} is isomorphic to U_S . Then by the Foiaş-Stratila theorem, T_A is isomorphic to S , and hence T_A is not LB. However, an induced automorphism from an LB automorphism is LB, a contradiction.

Another fruitful source of non-LB systems comes from taking Cartesian products of some natural LB systems. In Ornstein et al. (1982), it is proved that there exists a rank one (and hence LB) system whose Cartesian square is not LB. Moreover, in Ratner (1979), it was shown that the square of the horocycle flow is not LB (the horocycle flow itself being LB (Ratner 1978)). Recently, in Kanigowski and de la Rue (2019), the authors showed that there are staircase rank one transformations whose Cartesian product is not LB.

Rigid Sequences

Recall (see section “[Glossary and Notation](#)”) that an automorphism T of a standard probability Borel space $(X, \mathcal{B}, \mu)$ is called *rigid* if there exists a strictly increasing sequence $q_n \rightarrow \infty$ such that $\mu(T^{q_n} A \Delta A) \rightarrow 0$ as $n \rightarrow \infty$, for each $A \in \mathcal{B}$. (In fact, to have a global rigidity sequence, as

observed by Thouvenot, we only need to know that for each $A \in \mathcal{B}$ there is a sequence (q_n) so that $\mu(T^{q_n} A \Delta A) \rightarrow 0$. Equivalently, for each $f \in L^2(X, \mathcal{B}, \mu)$, $U_T^{q_n} f \rightarrow f$ in $L^2(X, \mathcal{B}, \mu)$ (it is not hard to see that the latter is equivalent that $\int_0^1 e^{2\pi i q_n x} d\sigma_f(x) \rightarrow 1$ for any σ_f representing the maximal spectral type of T , $\|f\| = 1$). We call (q_n) a *rigidity sequence* of T . Rigidity is one of (purely spectral) the fundamental phenomena in ergodic theory. Assuming that T is aperiodic, it is not hard to see that for any rigidity sequence (q_n) , we must have $q_{n+1} - q_n \rightarrow \infty$. Typical automorphism is rigid and weakly mixing, but since weak mixing implies $U_T^n \rightarrow 0$ weakly on $L_0^2(X, \mathcal{B}, \mu)$ along a sequence of n of full density, there is no much “room” left for rigidity sequences. So positive density sequences cannot be rigid, but beyond that, in the class of zero density sequences, there can be other, for example algebraic in nature, obstructions for rigidity. For example, as noticed in Bergelson et al. (2014) and Eisner and Grivaux (2011), if $P \in \mathbb{Q}[x]$ is any non-zero polynomial taking integer values on $\mathbb{Z}$, then the sequence $(P(n))$ cannot be rigid for any ergodic automorphism. It is also easy to see that (2^n) is a rigidity sequence while $(2^n + 1)$ is not.

A systematic study of sequences which can be rigidity sequences was originated in Bergelson et al. (2014) and Eisner and Grivaux (2011). Both papers use harmonic analysis approach to construct rigid sequences (via the standard Gaussian functor). Other constructions are also presented in Bergelson et al. (2014), rank one constructions, weighted operators, and Poisson suspensions, while Eisner and Grivaux (2011) rather concentrates on so-called linear dynamical systems and studies rigidity for weakly mixing automorphisms. One of the results in Bergelson et al. (2014) and Eisner and Grivaux (2011) states that if either $q_{n+1}/q_n \rightarrow \infty$ or q_{n+1}/q_n is an integer, then (q_n) is a rigidity sequence (for a weakly mixing automorphism). On the other hand, Eisner and Grivaux in Eisner and Grivaux (2011) give an example of a rigid sequence (q_n) , for a weakly mixing automorphism, such that $q_{n+1}/q_n \rightarrow 1$. As a matter of fact, both Bergelson et al. (2014) and Eisner and Grivaux (2011) deal with the case of denominators (q_n) of an irrational $\alpha \in [0, 1)$

(which are obviously rigidity sequences for the corresponding irrational rotations $Tx = x + \alpha$ on the additive circle) to show that such sequences are rigid for some weakly mixing automorphisms. Let us also mention Aaronson’s result (Aaronson 1979): *Given any sequence (r_n) of density 0, there is a sequence (q_n) such that $q_n < r_n$, $n \geq 1$, and (q_n) is rigid for some weakly mixing automorphisms.*

Moreover, in Bergelson et al. (2014), two basic questions have been formulated: *Given any sequence rigid for some T with discrete spectrum, must it be rigid for some weakly mixing automorphism? What about the converse?*

The positive answer to the first question was given by Adams (2015) and Fayad and Thouvenot (2014). On the other hand, surprisingly, Fayad and Kanigowski (2015) answered negatively the second question: There are rigidity sequences (for weakly mixing automorphisms) which are not rigidity sequences for any rotation. For a strengthening of this result (the existence of a rigidity sequence which, as a subset, is dense in the Bohr topology on $\mathbb{Z}$), see Griesmer (2019).

One can also consider a notion stronger than rigidity, called IP-rigidity (see, e.g., Bergelson et al. (2014)): (q_n) is an IP-rigidity sequence for an automorphism T acting on $(X, \mathcal{B}, \mu)$ if $T^\xi \rightarrow Id$ in the strong topology of $L^2(X, \mathcal{B}, \mu)$ in the IP sense, that is, when $\xi \rightarrow \infty$, where $\xi = q_{m_1} + \dots + q_{m_k}$ and we require that the smallest element q_{m_1} is going to ∞ . This notion is studied in Aaronson et al. (2014) relating it to non-singular ergodic theory (more precisely, to groups of so-called L^∞ -eigenvalues of non-singular automorphisms). As proved in Aaronson et al. (2014), in this category, the answer to the second question (above) from Bergelson et al. (2014) turns out to be positive. Moreover, the paper provides an example of a super-lacunary sequence (which must be a rigidity sequence by Bergelson et al. (2014) and Eisner and Grivaux (2011)) which is not an IP-rigid.

In the recent preprint (Badea et al. 2018), rigidity sequences are compared to other classical notions in harmonic analysis. It is proved that rigidity sequences (q_n) are nullpotent, i.e., there exists a topology τ on $\mathbb{Z}$ making it a topological

group such that $q_n \rightarrow 0$, but they are never Kazhdan. (A subset $B \subset \mathbb{Z}$ is called Kazhdan if there exists $\varepsilon > 0$ such that each unitary operator U on a separable Hilbert space H having a unit vector x with $\sup_{n \in B} \|U^n x - x\| < \varepsilon$ has a non-zero fixed point.) We find also there a rather surprising result that the family of all rigidity sequences considered as a subset in $\mathbb{Z}^{\mathbb{N}}$ is Borel.

Spectral Theory of Parabolic Dynamical Systems

We say a system is *algebraic* if it is a $\mathbb{Z}$ (or $\mathbb{R}$) translation on a quotient of a Lie group by a lattice. Spectral theory (and mixing properties) of *algebraic systems* is by now well understood. The two main classes are actions on quotients of semi-simple and nilpotent Lie groups. In the first case, the two main examples are horocycle and geodesic flows on quotients of $SL(2, \mathbb{R})$. More generally, one can talk about quasi-unipotent and partially hyperbolic actions. Recall that in the setting of algebraic actions, being quasi-unipotent is equivalent to *zero entropy*, while being partially hyperbolic is equivalent to *positive entropy*.

It is known that in both cases, the spectrum is countable Lebesgue (we refer the reader to Katok and Thouvenot (2006) for a nice description of spectral theory of horocycle and geodesic flows). Actions on quotients of nilpotent Lie groups are also known to have countable Lebesgue spectrum in the orthocomplement of the eigenspace (we refer to Parry (1970) for details). Quantitative mixing (and higher-order mixing) of algebraic systems is also well understood (we refer the reader to a recent paper, Björklund et al. (2020), for general results on decay of correlations for algebraic systems on semi-simple Lie groups).

Much less is known in spectral theory of parabolic systems beyond algebraic world. There is no strict definition for a system to be parabolic. However, characteristic features of parabolic systems are zero entropy, polynomial orbit growth, strong mixing, and equidistribution properties. We describe some classes of (non-algebraic) parabolic systems below. One of the main difficulties in studying non-algebraic

systems is a lack of many tools from representation theory, which is available in the algebraic setting. Below, we focus on known results and questions in spectral theory of non-algebraic parabolic systems.

Time-Changes of Algebraic Systems

Perhaps the simplest class of non-algebraic parabolic dynamical systems is given by time-changes (or reparametrizations) of algebraic systems. As for algebraic systems, it is natural to consider separately the cases of time-changes of unipotent systems and nilpotent systems. We do it in two paragraphs below.

Time-changes of unipotent systems In recent years, we witnessed substantial development in understanding the theory of time-changes of unipotent flows. The first (and most studied) case is that of smooth time changes of horocycle flows. Recall first that M. Ratner (1987) established measures and joinings rigidity phenomena for time-changes of horocycle flows that are analogous to Ratner's theory in algebraic setting. In particular in Ratner (1986), Ratner proved the H -property for all (sufficiently smooth time-changes). It is not known if Ratner's joinings and measure rigidity also holds for time-changes of general unipotent flows.

Mixing for smooth time-changes of horocycle flows was established by Marcus (1977), generalizing earlier work of Kushnirenko (1974) who required additionally small derivative in the geodesic direction. A crucial result for the theory is by L. Flaminio and G. Forni (2003), where the authors classify all invariant distributions and as a consequence show that a typical time-change is not a (measurable) quasi-coboundary (and hence the time-change is not trivially isomorphic to the original flow). A. Katok and J.-P. Thouvenot conjectured (Katok and Thouvenot 2006) that every sufficiently smooth time change of the horocycle flow has countable Lebesgue spectrum. A partial answer to this conjecture was given by G. Forni and C. Ulcigrai (2012), where the authors show that the maximal spectral type of the time-changed flow remains Lebesgue (see also a result of R. Tiedra de Aldecoa (2012), where the absolute continuity of the spectrum is proven, and

Tiedra de Aldecoa (2015b, 2017) for further applications of the commutator method in ergodic theory). A full solution of the Katok-Thouvenot conjecture (i.e., countable Lebesgue spectrum) was recently given by G. Forni, B. Fayad, and A. Kanigowski (2016). Generalizing the approach from Forni and Ulcigrai (2012), L. Simonelli (2018) showed that the spectrum of smooth time-changes of general unipotent flows remains Lebesgue. It is not known if the multiplicity is infinite, but it seems that the approach from Fayad et al. (2016) has the potential of being applicable in this setting.

Recall that Ratner's work (Ratner 1983) allows one to classify joinings between horocycle flows. Recently, there was a progress in understanding joinings for time-changes of horocycle flows. In Kanigowski et al. (2018) (see also a result in Flaminio and Forni (2019)), the authors show that there is a strong dichotomy for two smooth time-changes of horocycle flows: Either the time-change functions are cohomologous or the resulting time-changed flows are disjoint.

Even though quantitative mixing for time-changes of unipotent flows is now well understood, not much is known for quantitative higher-order correlations:

Question 1 *Is the decay of higher correlations for non-trivial time-changes of horocycle flows (or more generally, unipotent flows) polynomial?*

For trivial time-changes, i.e., for the horocycle flow, decay of higher correlations is indeed polynomial by a recent result of M. Björklund, M. Einsiedler, and A. Gorodnik (2020) (in fact this applies to general unipotent flows).

Time-changes of nilpotent systems Recall that nilpotent flows are never weakly mixing since they always have a non-trivial Kronecker factor. An interesting question is therefore whether one can improve mixing properties of the system by a time-change. The first result in this direction by A. Avila, G. Forni, and C. Ulcigrai (2011) is that there exists a dense set of smooth functions on the *Heisenberg nilmanifold* such that the resulting time-changed Heisenberg flow is *mixing*. This result was

strengthened by D. Ravotti in Ravotti (2019) to quasi-Abelian nilflows and recently to all nilflows by Avila, Forni, Ravotti, and Ulcigrai (2019). It is important to mention that the mixing mechanism is nonquantitative. Therefore, two questions are natural to ask: What are mixing properties of general time-changes of nilflows and can one obtain some quantitative mixing results? The only case in which some progress has been recently made is that of time-changes of Heisenberg nilflows. In Forni and Kanigowski (2017), the authors show stretched polynomial decay of correlations for smooth time-changes of full measure set of Heisenberg nilflows (parameterized by the frequency of the Kronecker factor). In the case the flow is of *bounded type*, the authors prove polynomial speed of decay of correlations. Moreover, in Forni and Kanigowski (2020), the authors show that for time-changes of bounded type Heisenberg nilflows, every non-trivial time-change enjoys the *R*-property and as a consequence is *mildly mixing*. Moreover, in the above setting, it also follows that every mixing time-change is mixing of all orders.

Mixing and spectral properties of time-changes of general nilflows are poorly understood. In particular, the following question seems interesting:

Question 2 *Are all non-trivial smooth time-changes of general nilflows mixing?*

The mixing mechanism from Avila et al. (2011) (see also Ravotti (2019) and Avila et al. (2019)) is nonquantitative. Therefore, the following question seems to be far more challenging:

Question 3 *Does there exist a smooth time change of a general nilflow with AC (Lebesgue) spectrum?*

Special Flows, Flows on Surfaces, and Interval Exchange Transformations

In this section, we will describe spectral results for special flows over interval exchange transformations (IETs) (irrational rotations begin a particular case). As described below, such special flows arise as representations of smooth locally Hamiltonian flows on surfaces.

Interval Exchange Transformations

To define an interval exchange transformation (IET) of m intervals, we need a permutation π of $\{1, \dots, m\}$ and a probability vector $\lambda = (\lambda_1, \dots, \lambda_m)$ (with positive entries). Then, we define $T = T_{\lambda, \pi}$ of $[0, 1)$ by putting

$$T_{\lambda, \pi}(x) = x + \beta_i^\pi - \beta_i \text{ for } x \in [\beta_i, \beta_{i+1}),$$

where $\beta_i = \sum_{j < i} \lambda_j$, $\beta_i^\pi = \sum_{\pi j < \pi i} \lambda_j$. Obviously, each IET preserves Lebesgue measure. One of possible approaches to study ergodic properties of IET is an “a.e.” approach “seen” in the definition of $T_{\lambda, \pi}$. It is based on the fundamental fact that the induced transformation on a subinterval of $[0, 1)$ is also IET (see Cornfeld et al. (1982)). This leads to a very delicate and deep mathematics based on Rauzy induction, which is a way of inducing on special intervals, considering only irreducible permutations whose set is partitioned into orbits of some maps (any such an orbit is called a *Rauzy class*). If now $\mathcal{R}$ is a Rauzy class of permutations and λ lies in the standard simplex Δ_{m-1} , then the Rauzy induction together with a natural renormalization leads to a map $\mathcal{P} : \mathcal{R} \times \Delta_{m-1} \rightarrow \mathcal{R} \times \Delta_{m-1}$. For a better understanding of the dynamics of the Rauzy map, Veech (1982b) introduced the space of *zippered rectangles*. A zippered rectangle associated with the Rauzy class $\mathcal{R}$ is a quadruple (λ, h, a, π) , where $\lambda \in \mathbb{R}_+^m$, $h \in \mathbb{R}_+^m$, $a \in \mathbb{R}_+^m$, $\pi \in \mathcal{R}$ and the vectors h and a satisfy some equations and inequalities. Every zippered rectangle (λ, h, a, π) determines a Riemann structure on a compact connected surface. Denote by $\Omega(\mathcal{R})$ the space of all zippered rectangles, corresponding to a given Rauzy class $\mathcal{R}$ and satisfying the condition $\langle \lambda, h \rangle = 1$. In Veech (1982b), Veech defined a form $(P^t)_{t \in \mathbb{R}}$ on the space $\Omega(\mathcal{R})$ putting

$$P^t(\lambda, h, a, \pi) = (e^t \lambda, e^{-t} h, e^{-t} a, \pi)$$

and extended the Rauzy map. On so-called Veech moduli space of zippered rectangles, the flow (P^t) becomes the *Teichmüller flow*, and it preserves a natural Lebesgue measure class; by passing to a transversal and projecting the measure on the

space of IETs $\mathcal{R} \times \Delta_{m-1}$, Veech has proved the following fundamental theorem (Veech 1982b; see also Masur (1982)) which is a generalization of the fact that Gauss measure $\frac{1}{\ln 2} \frac{1}{1+x} dx$ is invariant for the Gauss map which sends $t \in (0, 1)$ into the fractional part of its inverse.

Theorem 10 (Veech and Masur) (1982) *Assume that $\mathcal{R}$ is a Rauzy class. There exists a σ -finite measure $\mu_{\mathcal{R}}$ on $\mathcal{R} \times \Delta_{m-1}$ which is $\mathcal{P}$ -invariant, equivalent to “Lebesgue” measure, conservative and ergodic.*

In Veech (1982b), it is proved that a.e. (in the above sense) IET is then of rank one (and hence is ergodic and has a simple spectrum). He also formulated as an open problem whether we can achieve the weak mixing property a.e. This has been answered in positive by A. Avila and G. Forni (2007) (for π which is not a rotation).

Katok (1980) proved that IET have no mixing factors (in fact his proof shows more: The IET class is disjoint from the class of mixing transformations). By their nature, IET transformations are of finite rank (see Cornfeld et al. (1982)), so they are of finite multiplicity. They need not be of simple spectrum (see remarks in Katok and Thouvenot (2006), pp. 88–90). It remains an open question whether an IET can have a non-singular spectrum. Joining properties in the class of exchange of 3 and more intervals are studied in Ferenczi et al. (2004, 2005). In Chaika and Eskin (2018), the authors show that a.e. 3-IET is not simple. This answers a special case of a question of Veech (1982a) whether a.e. IET is simple. The case of d -IET’s with $d > 3$ is still widely open.

Smooth Flows on Surfaces and Their Special Representations

We consider a closed, connected, smooth, and orientable surface S of genus $g \geq 1$. Let $X : S \mapsto TS$ be a smooth vector field with finitely many fixed points and such that the corresponding flow (ϕ_t^X) preserves a smooth area form ω . The form (ϕ_t^X) is called a *locally Hamiltonian flow*; it is locally given by a smooth Hamiltonian H (up to an additive constant), so that (ϕ_t^X) is a solution to

$$\frac{dx}{dt} = \frac{\partial H}{\partial y}, \quad \frac{dy}{dt} = -\frac{\partial H}{\partial x}. \quad (3)$$

There has recently been some progress in understanding ergodic and spectral properties of locally Hamiltonian flows. It follows by Arnold (1991) that (ϕ_t^X) can be decomposed into finite number of topological discs filled with periodic orbits and finite number of connected components on which the flow is minimal. We are always interested in properties of (ϕ_t^X) on the minimal components. A classical way of studying locally Hamiltonian flows is through their *special representations*. On each minimal component, one chooses a smooth transversal and represents (ϕ_t^X) by the first return map (the base transformation T) and the first return time (the roof function f). We will denote the special representation of (ϕ_t^X) by T^f . It is well-known that T is either a rotation (in genus 1 case) or an interval exchange transformation (in higher genus) and the roof function is smooth except for finitely many points at which it explodes according to the nature of the fixed points. Notice first that if (ϕ_t^X) has no fixed points, then S is a two-dimensional torus (by Gauss-Bonnet). In this case, (ϕ_t^X) is a smooth time-change of a linear flow on the two-torus. Therefore, it is never mixing (Kočergin 1972) and spectrally disjoint from mixing flows (Frączek and Lemańczyk 2004) and in particular has purely singular spectrum. From now on, we will therefore assume that the set of fixed points is non-empty. If all fixed points are non-degenerated (i.e., Hessian of H is non-zero at every critical point), the roof function f has *logarithmic singularities*, i.e., it blows up logarithmically at finitely many points. Logarithmic singularities can be either *symmetric* (for instance, if there are no saddle connections) or *asymmetric* (in case there is a saddle loop). Ergodic properties of (ϕ_t^X) depend strongly on whether the roof has symmetric or asymmetric singularities. If there are *degenerated* fixed points, Kočergin (Kočergin 1975) constructed smooth Hamiltonian for which the roof function at singularities blows up power-like (like x^γ , $\gamma \in (-1, 0)$). By changing a speed as it is done in Frączek and Lemańczyk (2006) so

that critical points of the vector field in (3) become singular points, Arnold's special representation is transformed to a special flow over the same irrational rotation however under a piecewise smooth function. If the sum of jumps is not zero, then in fact we come back to von Neumann's class of special flows considered in von Neumann (1932). Similar classes of special flows (when the roof function is of bounded variation) are obtained from ergodic components of flows associated with billiards in convex polygons with rational angles (Katok and Zemlyakov 1975).

It is convenient to express ergodic and spectral properties of locally Hamiltonian flows in terms of their special representation. We will do it in the next subsection.

Special Flows over Rotations and Interval Exchange Transformations

Special flows were introduced to ergodic theory by von Neumann in his fundamental work (von Neumann 1932) in 1932. Also in that work he explains how to compute eigenvalues for special flows, namely: $r \in \mathbb{R}$ is an eigenvalue of T^f if and only if the following functional equation

$$e^{2\pi i r f(x)} = \zeta(x) / \zeta(Tx)$$

has a measurable solution $\zeta : X \rightarrow \mathbb{T}$. We recall also that the classical Ambrose-Kakutani theorem asserts that practically each ergodic flow has a special representation (Cornfeld et al. 1982; see also Rudolph's theorem on special representation therein). As described in the previous subsection, we will consider the case when the roof functions is of bounded variation, has logarithmic singularities (symmetric or asymmetric), or has power singularities.

Bounded Variation Roof Function

Kočergin (1972) showed that special flows over irrational rotations and under bounded variation functions are never mixing. This has been strengthened in Frączek and Lemańczyk (2004) to the following: *If T is an irrational rotation and f is of bounded variation, then the special flow T^f is spectrally disjoint from all mixing flows.* In

particular, all such flows have singular spectra. Moreover, in Frączek and Lemańczyk (2004), it is proved that whenever the Fourier transform of the roof function f is of order $O(\frac{1}{n})$, then T^f is disjoint from all mixing flows (see also Frączek and Lemańczyk (2003)). In fact in the papers (Frączek and Lemańczyk 2003, 2004, 2005, 2006), the authors discuss the problem of disjointness of those special flows with all ELF-flows conjecturing that no flow of probabilistic origin has a smooth realization on a surface. In Lemańczyk and Wysokińska (2007), the analytic case is considered, leading to a “generic” result on disjointness with the ELF class generalizing the classical Shklover’s result on the weak mixing property (Shklover 1967). A. Katok (1980) proved the absence of mixing for special flows over IET when the roof function is of bounded variation (see also Ryzhikov (1994a)). Katok’s theorem was strengthened in Frączek and Lemańczyk (2005) to the disjointness theorem with the class of mixing flows. A. Avila and G. Forni (Avila and Forni 2007) proved that a.e. translation flow on a surface (of genus at least two) is weakly mixing (which is a drastic difference with the linear flow case of the torus, where the spectrum is always discrete).

One important property in the class of special flows over rotations and IETs is Ratner’s property (R-property). This property may be viewed as a particular way of divergence of orbits of close points; it was shown to hold for horocycle flows by M. Ratner (1983). We refer the reader to Ratner (1983) and the survey article in Thouvenot (1995) for the formal definitions and basic consequences of R-property. In particular, R-property implies “rigidity” of joinings and it also implies the PID property; hence, mixing and R-property imply mixing of all orders. In Frączek and Lemańczyk (2006) and Frączek et al. (2007), a version of R-property is shown for the class of von Neumann special flows (however, α is assumed to have bounded partial quotients). This allowed one to prove there that such flows are even mildly mixing (mixing is excluded by a Kochergin’s result). The eigenvalue problem (mainly how

many frequencies can have the group of eigenvalues) for special flows over irrational rotations is studied in Fayad et al. (2001), Fayad and Windsor (2007), and Guenais and Parreau (2005).

It follows from Frączek and Lemańczyk (2004) that von Neumann’s flows have singular spectrum. However, nothing is known about their multiplicity.

Question 4 *What is the spectral multiplicity of von Neumann flows?*

Symmetric Logarithmic Singularities

Kochergin (Kochergin 1976a) proved the absence of mixing for flows, where the roof function has finitely many singularities; however, some Diophantine restriction is put on α . In Lemańczyk (2000), where also the absence of mixing is considered for the symmetric logarithmic case, it was conjectured (and proved for arbitrary rotation) that a necessary condition for mixing of a special flow T^f (with arbitrary T and f) is the condition that the sequence of distributions $\left(\left(f_0^{(n)} \right)_* \right)_n$ tends to δ_∞ in the space of probability measures on $\overline{\mathbb{R}}$. K. Schmidt (Schmidt 2002) proved it using the theory of cocycles and extending a result from Aaronson and Weiss (2000) on tightness of cocycles. Ulcigrai (2011) showed that for every $d \geq 2$ and a.e. IET of d intervals, the corresponding special flow T^f is *not* mixing. However, Chaika and Wright (2019) proved existence of an IET T such that T^f is mixing. Notice that by Frączek and Lemańczyk (2004), it also follows that for a.e. irrational rotation T^f has purely singular spectral type. Recent result (Chaika et al. 2019) shows that one can also prove singularity of the spectrum for *symmetric IETs* in the base. This in particular shows that minimal flows on genus 2 surfaces (with two isometric saddles) have purely singular spectral type. In Kanigowski and Kułaga-Przymus (2016), the authors showed that if T is an IET of *bounded type*, then T^f is mildly mixing (and has the R-property). For symmetric logarithmic singularities, the following two questions are open:

Question 5 *What is the maximal spectral type of T^f for a general IETs?*

Nothing is known about multiplicity of the spectrum in this setting.

Asymmetric Logarithmic Singularities

In this case, mixing properties are different. Khanin and Sinai (1992) showed that T^f is mixing for a.e. irrational rotation T . This was strengthened by Ulcigrai to a.e. IET (Ulcigrai 2007). Moreover (Ravotti 2017), Ravotti obtained quantitative mixing estimates (with sub-logarithmic speed of decay of correlations). Not much was known about multiple mixing for T^f . This changed recently: In Fayad and Kanigowski (2016), the authors showed that for a.e. irrational rotation, T^f enjoys a variant of the R -property and hence is multiple mixing. This was strengthened to a.e. IETs in Kanigowski et al. (2019). The following two questions seem to be natural (the first one already stated as Questions 34, 35 in Fayad and Krikorian (2018)):

Question 6 *What is the maximal spectral type of T^f ?*

Moreover, one can ask about quantitative higher-order decay:

Question 7 *Is the decay of higher-order correlations sub-logarithmic?*

Both of the above questions are open even for rotations in the base. As in the symmetric case, nothing is known about multiplicity of the spectrum.

Power Singularities

In case f has power singularities, T^f was shown to be mixing by Kochergin (1975) (for any irrational rotation and IETs). In Fayad and Kanigowski (2016), the authors show that if T is a rotation of bounded type, then T^f is multiple mixing (it enjoys a variant of the R -property). Polynomial decay for some flows with power singularities was

obtained by B. Fayad (2002b). In Fayad et al. (2016), the authors considered the spectrum of T^f . They showed that if f has sufficiently strong power singularity (of the form $x^{-1+\eta}$ for small $\eta > 0$), then T^f has countable Lebesgue spectrum for a.e. irrational rotation. To the best of our knowledge, this is the only result dealing with multiplicity for smooth surface flows. The following problems are natural (see Question 34 in Fayad and Krikorian (2018)):

Question 8 *What is the maximal spectral type of T^f when f has power singularities? What is the multiplicity?*

To answer this, one needs to consider general IETs in the base, as well as functions with weaker power singularity than in Fayad et al. (2016). The following question is still open (Question 38 in Fayad and Krikorian (2018)):

Question 9 *Are all mixing surface flows mixing of all orders?*

Finally, it may also be useful to show that smooth flows on surfaces are disjoint from flows of probabilistic origin – see del Junco and Lemańczyk (1999, 2005), Lemańczyk et al. (2011), Ryzhikov and Thouvenot (2006), and Thouvenot (2000).

B. Fayad (2006) gives a criterion that implies singularity of the maximal spectral type for a dynamical system on a Riemannian manifold. As an application, he gives a class of smooth mixing flows (with singular spectra) on $\mathbb{T}^3$ obtained from linear flows by a time change (again this is a drastic difference with dimension two, where a smooth time change of a linear flow leads to non-mixing flows (Cornfeld et al. 1982)).

We mention at the end that if we drop here (and in other problems) the assumption of regularity of f , then the answers will be always positive because of the LB theory; in particular, there is a section of any horocycle flow (it has the LB property (Ratner 1978)) such that in the corresponding special representation T^f , the map T is an irrational rotation.

Using a Kochergin's result (Kochergin 1976b) on cohomology (see also Katok (2003a) and Rudolph (1986)), the L^1 -function f is cohomologous to a positive function g which is even continuous; thus, T^f is isomorphic to T^g .

Spectral Theory for Locally Compact Groups of Type I

This section has been written by A. Danilenko.

Groups of Type I

The spectral theory presented here for Abelian group actions extends potentially to probability-preserving actions of non-Abelian locally compact groups of type I. We now provide the definition of type I. Let G be a locally compact second countable group, $\mathcal{H}$ a separable Hilbert space, and $\pi : G \ni g \mapsto \pi(g)$ a (weakly) continuous unitary representation of G in $\mathcal{H}$. We say that π is of type I if there is a subset $A \subset \{1, 2, \dots, +\infty\}$ such that π is unitarily equivalent to the orthogonal sum $\bigoplus_{k \in A} U_k \otimes I_k$, where U_k is a unitary representation of G with a simple spectrum and I_k is the trivial representation in the Hilbert space of dimension k .

Definition 1 If every unitary representation of G is of type I, then G is called of type I.

Denote by $\widehat{G}$ the *unitary dual* of G , i.e., the set of unitarily equivalent classes of all irreducible unitary representations of G . If G is Abelian, then every irreducible representation of G is one dimensional. Hence, $\widehat{G}$ is identified naturally with the group of characters of G . In the general case, let $\text{Irr}_n(G)$ stand for the set of all irreducible unitary representations of G in the n -dimensional separable Hilbert space $\mathcal{K}_n$, $1 \leq n \leq +\infty$. Endow it with the natural Borel structure, i.e., the smallest one in which the mapping $\pi \mapsto \langle \pi(g)f, h \rangle$ is Borel for every $g \in G$ and $f, h \in \mathcal{K}_n$. It is standard. Let $\widehat{G}_n$ be the quotient of $\text{Irr}_n(G)$ by the unitary equivalence relation. Endow $\widehat{G}_n$ with the quotient Borel structure. Since $\widehat{G}$

$= \bigsqcup_{n=1}^{\infty} \widehat{G}_n \bigsqcup \widehat{G}_{\infty}$, we obtain a Borel structure on $\widehat{G}$. It is called *Mackey-Borel structure* on $\widehat{G}$. By the Glimm theorem, the Mackey-Borel structure is standard if and only if G is of type I (Glimm 1960).

For $n \in \mathbb{N} \cup \{\infty\}$, denote by I_n the identity operator on $\mathcal{K}_n$. Then for each unitary representation π of G in $\mathcal{H}$, there are a measure λ , a measurable field $\widehat{G} \ni \omega \mapsto \mathcal{H}_{\omega}$ of Hilbert spaces, and a measurable field $\widehat{G} \ni \omega \mapsto V_{\omega}$ of irreducible unitary G -representations such that $V_{\omega} \in \omega$, and $\mathcal{H}_{\omega}$ is the space of V_{ω} on $\widehat{G}$ and a measurable map $m : \widehat{G} \rightarrow \mathbb{N} \cup \{+\infty\}$ such that

$$\mathcal{H} = \int_{\widehat{G}}^{\oplus} \mathcal{H}_{\omega} \otimes \mathcal{K}_{m(\omega)} d\lambda(\omega) \text{ and}$$

$$\pi(g) = \int_{\widehat{G}}^{\oplus} V_{\omega}(g) \otimes I_{m(\omega)} d\lambda(\omega).$$

It appears that if G is of type I, then the equivalence class of λ is defined uniquely by π , and the function m is defined up to a λ -zero subset. We call the class of λ the *maximal spectral type* of π , and we call m the *spectral multiplicity* of π . If we have a probability-preserving action $T = (T_g)_{g \in G}$ of G , then we can consider the corresponding Koopman unitary representation π of G . The maximal spectral type of π and the spectral multiplicity of π are called the *maximal spectral type* of T and the *spectral multiplicity* of T , respectively. If G is Abelian, then these concepts coincide with their classic counterparts considered above in the survey.

All compact groups, Abelian groups, connected semi-simple Lie groups, and nilpotent Lie groups (or, more generally, exponential Lie groups) are of type I. Each subgroup of $GL(n, \mathbb{R})$ determined by a system of algebraic equations is also of type I. Solvable Lie groups can be as of type I as not of type I. If G is a countable (discrete) group, then G is of type I if, and only if it is virtually Abelian, i.e., it contains an Abelian subgroup of finite index (Thoma 1964).

Even if we know that a non-Abelian group G is of type I, it is usually not an easy problem to

describe $\widehat{G}$ explicitly. Kirillov introduced an *orbit method* for a description of $\widehat{G}$ when G is a connected, simply connected nilpotent Lie group (the method was developed further for solvable groups). He identified $\widehat{G}$ with the space of orbits for the co-adjoint G -action on the dual space $\mathfrak{g}^*$ of its Lie algebra $\mathfrak{g}$. Though Kirillov's method gives an algorithm how to describe $\widehat{G}$, not so many groups are known for which the unitary dual is described explicitly.

Spectral Properties of Heisenberg Group Actions

The three-dimensional Heisenberg group $H_3(\mathbb{R})$ is perhaps one of the simplest examples of non-Abelian nilpotent Lie groups for which the orbit method leads to a very concrete description of the unitary dual (Kirillov 2004). Recall that

$$H_3(\mathbb{R}) = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} : a, b, c \in \mathbb{R} \right\}.$$

For simplicity, we will denote the matrix $\begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}$ by $[a, b, c]$. The unitary dual $\widehat{H_3(\mathbb{R})}$ is identified with $\mathbb{R}^2 \sqcup \mathbb{R}^*$ endowed with the natural standard Borel structure. Every irreducible unitary representation of $H_3(\mathbb{R})$ is unitarily equivalent to either a one-dimensional $\pi_{\alpha, \beta}$ with $(\alpha, \beta) \in \mathbb{R}^2$ or an infinite dimensional π_γ in $L^2(\mathbb{R}, \text{Leb})$, with $\gamma \in \mathbb{R}^* := \mathbb{R} \setminus \{0\}$, such that

$$\pi_{\alpha, \beta}[a, b, c] := e^{2\pi i(\alpha a + \beta b)},$$

$$\pi_\gamma[a, b, c]f(x) := e^{2\pi i(c + bx)}f(x + a), \quad f \in L^2(\mathbb{R}, \text{Leb}).$$

Now, given a measure-preserving $H_3(\mathbb{R})$ -action on a standard probability space (X, μ) , let U denote the corresponding Koopman representation in $L^2(X, \mu)$. Then there are a probability measure $\sigma^{1,2}$ on $\mathbb{R}^2$, a function $l^{1,2} : \mathbb{R}^2 \rightarrow \mathbb{N}$ a probability measure σ^3 on $\mathbb{R}^*$, a function $l^3 : \mathbb{R}^* \rightarrow \mathbb{N}$ such that

$$L^2(X, \mu) = \int_{\mathbb{R}^2}^{\oplus} \bigoplus_{j=1}^{l^{1,2}(\alpha, \beta)} \mathbb{C} d\sigma^{1,2}(\alpha, \beta) \oplus$$

$$\int_{\mathbb{R}^*}^{\oplus} \bigoplus_{j=1}^{l^3(\alpha, \beta)} L^2(\mathbb{R}, \text{Leb}) d\sigma^3(\gamma) \text{ and}$$

$$U = \int_{\mathbb{R}^2}^{\oplus} \bigoplus_{j=1}^{l^{1,2}(\alpha, \beta)} \pi_{\alpha, \beta} d\sigma^{1,2}(\alpha, \beta) \oplus$$

$$\int_{\mathbb{R}^*}^{\oplus} \bigoplus_{j=1}^{l^3(\alpha, \beta)} \pi_\gamma d\sigma^3(\gamma).$$

We now compare the spectral properties of T with the spectral properties of the restriction T to the center of $H_3(\mathbb{R})$. The center is the subgroup $\{[0, 0, t] : t \in \mathbb{R}\}$.

Proposition 1 (Danilenko (2014)) *The maximal spectral type of the flow $(T_{[0,0,t]})_{t \in \mathbb{R}}$ contains the measure $\sigma^{1,2}(\mathbb{R}^2)\delta_0 + \sigma^3$. The corresponding spectral multiplicity is $\int_{\mathbb{R}^2} l^{1,2} d\sigma^{1,2}$ at the point $t = 0$ and the infinity if $t \in \mathbb{R}^*$.*

Theorem 11 (Danilenko (2014)) *If $(T_{[0,0,t]})_{t \in \mathbb{R}}$ is ergodic then:*

1. $\sigma^{1,2}(\mathbb{R}^2 \setminus \{(0, 0)\}) = 0$, i.e., there are no non-trivial one-dimensional representations in the spectral decomposition of U . The maximal spectral type of T equals the maximal spectral type of the restriction of T to the center of $H_3(\mathbb{R})$ (modulo the natural identification);
2. T is mixing (see also Ryzhikov (1994b, 1996));
3. the weak closure of the group $\{T_g : g \in H_3(\mathbb{R})\}$ in $\text{Aut}(X, \mu)$ is the union of $\{T_g : g \in H_3(\mathbb{R})\}$ and the weak closure of $\{T_{[0,0,t]} : t \in \mathbb{R}\}$;
4. if T is rigid, then $(T_{[0,0,t]})_{t \in \mathbb{R}}$ is rigid.

In Danilenko (2014), there were constructed explicit examples of mixing of all orders rank one (and hence zero entropy) actions T of the Heisenberg group.

The concept of simplicity for ergodic $H_3(\mathbb{R})$ -actions is defined in a similar way as for the Abelian actions. A simple $H_3(\mathbb{R})$ -action T has MSJ if the centralizer of the action is $\{T_{[0,0,t]} : t \in \mathbb{R}\}$. It was shown in Danilenko (2014) that the examples of mixing $H_3(\mathbb{R})$ -actions constructed there satisfy also the following:

1. The flow $(T_{[0,0,t)})_{t \in \mathbb{R}}$ is simple and the centralizer of it equals the group $\{T_g : g \in H_3(\mathbb{R})\}$.
2. The transformation $T_{[0,0,1]}$ is simple and the centralizer of it equals $\{T_g : g \in H_3(\mathbb{R})\}$.
3. T has MSJ.

As a corollary, we obtain examples of mixing Poisson and mixing Gaussian (probability preserving) actions of $H_3(\mathbb{R})$.

Heisenberg Odometers

In Danilenko and Lemańczyk (2016), the authors isolated a special class of ergodic $H_3(\mathbb{R})$ -actions, called *the odometer* actions. Namely, let $\Gamma_1 \supset \Gamma_2 \supset \dots$ be a sequence of lattices in $H_3(\mathbb{R})$. Then, we can associate a sequence of homogeneous $H_3(\mathbb{R})$ -spaces intertwined with $H_3(\mathbb{R})$ -equivariant maps:

$$H_3(\mathbb{R})/\Gamma_1 \leftarrow H_3(\mathbb{R})/\Gamma_2 \leftarrow \dots$$

Denote by X the projective limit of this sequence. Then X is a compact Polish G -space. Endow each space $H_3(\mathbb{R})/\Gamma_n$ with the Haar probability measure. The projective limit of the sequence of these measures is a $H_3(\mathbb{R})$ -invariant probability measure μ . Of course, the $H_3(\mathbb{R})$ -action on (X, μ) is ergodic. It is called *the Heisenberg odometer associated with $(\Gamma_n)_{n=1}^\infty$* . We consider the Heisenberg odometers as non-commutative counterparts of the ergodic $\mathbb{Z}$ -actions and $\mathbb{R}$ -actions with pure point rational spectrum.

A complete spectral decomposition of the Heisenberg odometers is found in Danilenko and Lemańczyk (2016). Denote by $p : H_3(\mathbb{R}) \rightarrow \mathbb{R}^2$ the homomorphism $[a, b, c] \mapsto (a, b)$. The kernel of this homomorphism is the center of the Heisenberg group. Given a lattice Γ in $H_3(\mathbb{R})$, we denote by ξ_Γ a positive real such that $\Gamma \cap \text{Ker } p = \{[0, 0, n\xi_\Gamma] : n \in \mathbb{Z}\}$. Of course, $p(\Gamma)$ is a lattice in $\mathbb{R}^2$. If $p(\Gamma) = A(\mathbb{Z}^2)$ for some matrix $A \in GL(2, \mathbb{R})$, then we denote by $p(\Gamma)^*$ the dual lattice $(A^*)^{-1}\mathbb{Z}^2$ in $\mathbb{R}^2$.

Theorem 12 *Let U stand for the Koopman unitary representation of the Heisenberg odometer associated with a sequence of lattices $\Gamma_1 \supset \Gamma_2 \supset \dots$. If $\bigcup_{n=1}^\infty p(\Gamma_n)^*$ is not closed in $\mathbb{R}^2$, then*

$$U = \bigoplus_{(\alpha, \beta) \in \bigcup_{n=1}^\infty p(\Gamma_n)^*} \pi_{\alpha, \beta} \oplus \bigoplus_{0 \neq \gamma \in \bigcup_{n=1}^\infty \xi_{\Gamma_n}^{-1} \mathbb{Z}} \bigoplus_1^\infty \pi_\gamma.$$

An analogous decomposition is found also for the case where $\bigcup_{n=1}^\infty p(\Gamma_n)^*$ is closed in $\mathbb{Z}^2$ (see Danilenko and Lemańczyk (2016) for details). Thus, we see that the maximal spectral type of Heisenberg odometers is purely atomic.

For a decreasing sequence $\Gamma = (\Gamma_n)_{n=1}^\infty$ of lattices in $H_3(\mathbb{R})$, we let $S(\Gamma) := \bigcup_{n=1}^\infty p(\Gamma_n)^*$ and $\xi_\Gamma := \bigcup_{n=1}^\infty \xi_{\Gamma_n}^{-1} \mathbb{Z}$. The following theorem (except for the first claim) and the below remarks demonstrate a drastic difference between $H_3(\mathbb{R})$ -odometers and $\mathbb{Z}$ -odometers.

Theorem 13 *Two Heisenberg odometers T and T' associated with decreasing sequences of lattices Γ and Γ' respectively are unitarily equivalent if and only if $S(\Gamma) = S(\Gamma')$ and $\xi_\Gamma = \xi_{\Gamma'}$. The direct product $T \times T' := \left(T_g \times T'_g\right)_{g \in G}$ is not spectrally equivalent to any Heisenberg odometer. $T \times T'$ is ergodic if and only if $S(\Gamma) \cap S(\Gamma') = \{0\}$. $T \times T'$ is ergodic and has discrete maximal spectral type if and only if $S(\Gamma) \cap S(\Gamma') = \{0\}$ and $\xi_\Gamma \cap \xi_{\Gamma'} = \{0\}$.*

It was shown in Danilenko and Lemańczyk (2016) that Heisenberg odometers are not *iso-spectral*, i.e., the unitary equivalence, in general, does not imply isomorphism for the underlying $H_3(\mathbb{R})$ -actions. It was also shown in Danilenko and Lemańczyk (2016) that Heisenberg odometers are not *spectrally determined*: A Heisenberg odometer is constructed which is unitarily equivalent to an $H_3(\mathbb{R})$ -action which is not isomorphic to any Heisenberg odometer.

On the “Finitely Dimensional” Part of the Spectrum

Suppose that G is an arbitrary locally compact second countable group. If G is not of type I, then $\widehat{G}$ furnished with the Mackey-Borel structure is a “bad” (not standard) Borel space. Then a

decomposition of a unitary representation of G into irreducibles can be done in essentially nonunique way (see Kirillov (2004)). Nevertheless, this “badness” is related only to the infinite-dimensional part of the spectrum. Thus, the union $\bigsqcup_{n=1}^{\infty} \widehat{G}_n = \widehat{G} \setminus \widehat{G}_{\infty}$ is a “good” standard Borel space. Thus, given a measure-preserving G -action, we can study “the finitely dimensional part” of the spectrum of the corresponding Koopman unitary representation of G .

This approach was used by Mackey in Mackey (1964), where he made an attempt to extend the theory of actions with pure discrete spectrum to non-Abelian groups. For that, he isolated a class of ergodic G -actions T for which the Koopman representation U_T decomposes into a (countable) family of finite-dimensional irreducible representations. We note that the family is uniquely defined by T . Mackey called such T an *action with pure point spectrum*. He established a structure for these actions. He showed that T has a pure point spectrum if and only if T is isomorphic to a G -action by rotations on a homogeneous space of G by a compact subgroup. However, in general, in contrast with the Abelian case, the G -actions with pure point spectrum are not necessarily isospectral even in the case of finite G (see (Todd 1950; Lemańczyk et al. 2002) Section “Rokhlin Cocycles” for counterexamples).

In Lightwood et al. (2014), Lightwood, Şahin, and Ugarcovici considered certain odometer actions of the discrete Heisenberg group $H_3(\mathbb{Z}) := \{[a, b, c] : a, b, c \in \mathbb{Z}\}$. This group is not of type I (in contrast with $H_3(\mathbb{R})$) because each subgroup of finite index in $H_3(\mathbb{Z})$ is non-Abelian. Hence, $\widehat{H_3(\mathbb{Z})}$ is a “bad” Borel space. On the other hand, the (standard Borel) subspace $\bigsqcup_{n=1}^{\infty} \widehat{H_3(\mathbb{Z})}_n$ of it is explicitly described in Indukaev (2007). Consider now an $H_3(\mathbb{Z})$ -odometer T generated by a decreasing sequence $\Gamma_1 \supset \Gamma_2 \supset \dots$ of normal subgroups of finite index in $H_3(\mathbb{Z})$. We call such an odometer *normal*. Then, T has a pure point spectrum in the sense of Mackey. Moreover, the full list of the finitely dimensional irreducible unitary representations that occur in U_T is found in Lightwood et al. (2014) in terms of the sequence $(\Gamma_n)_{n=1}^{\infty}$. It was

shown later in Danilenko and Lemańczyk (2016) that the multiplicity of each irreducible component in U_T equals the dimension of this component. It was also proved in Danilenko and Lemańczyk (2016) that the normal $H_3(\mathbb{Z})$ -odometers are isospectral. Thus, the unitary equivalence of the Koopman representations implies isomorphism of the underlying normal $H_3(\mathbb{Z})$ -odometers.

Future Directions

We have already seen several cases, where spectral properties interact with measure-theoretic properties of a system. Let us mention a few more cases which require further research and deeper understanding.

We recall that the weak mixing property can be understood as a property complementary to discrete spectrum (more precisely to the distality (Furstenberg 1981)), or similarly mild mixing property is complementary to rigidity. This can be phrased quite precisely by saying that T is not weakly (mildly) mixing if and only if it has a non-trivial factor with discrete spectrum (it has a non-trivial rigid factor). It has been a question for quite a long time if in a sense mixing can be “built” on the same principle. In other words, we seek a certain “highly” non-mixing factor. It was quite surprising when in 2005 F. Parreau (private communication) gave the positive answer to this problem.

Theorem 14 (Parreau) *Assume that T is an ergodic automorphism of a standard probability Borel space $(X, \mathcal{B}, \mu)$. Assume moreover that T is not mixing. Then there exists a non-trivial factor (see below) of T which is disjoint from all mixing automorphisms.*

In fact, Parreau proved that each factor of T given by $\mathcal{B}_{\infty}(\rho)$ (this σ -algebra is described in Lemańczyk et al. (2000)), where $U_T^{n_k} \rightarrow J_{\rho}$, is disjoint from all mixing transformations. This proof leads to some other results of the same type, for example: *Assume that T is an ergodic automorphism of a standard probability Borel*

space. Assume that there exists a non-trivial automorphism S with a singular spectrum which is not disjoint from T . Then T has a non-trivial factor which is disjoint from any automorphism with a Lebesgue spectrum.

The problem of spectral multiplicity of Cartesian products for “typical” transformation studied by Katok (2003a) and then its solution in Ageev (2008) which we already considered in section “The Multiplicity Function” lead to a study of those T for which

$$(CS) \sigma^{(m)} \perp \sigma^{(n)} \text{ whenever } m \neq n,$$

where $\sigma = \sigma_T$ just stands for the reduced maximal spectral type of U_T (which is constantly assumed to be a continuous measure); see also Stepin’s article (Stepin 1986).

Usefulness of the above property (CS) in ergodic theory was already shown in del Junco and Lemańczyk (1992), where a spectral counterexample machinery was presented using the following observation: *If $\mathcal{A}$ is a T^{∞} -invariant sub- σ -algebra such that the maximal spectral type on $L^2(\mathcal{A})$ is absolutely continuous with respect to σ_T , then $\mathcal{A}$ is contained in one of the coordinate sub- σ -algebras $\mathcal{B}$.* Based on that in del Junco and Lemańczyk (1992), it is shown how to construct two weakly isomorphic actions which are not isomorphic or how to construct two non-disjoint automorphisms which have no common non-trivial factors (such constructions were previously known for so-called minimal self-joining automorphisms (Rudolph 1979)). See also Tikhonov (2002) for extensions of those results to $\mathbb{Z}^d$ -actions.

Prikhodko and Ryzhikov (2000) proved that the classical Chacon transformation enjoys the (CS) property. The SCS property defined in section “Glossary and Notation” is stronger than the (CS) condition above; the SCS property implies that the corresponding Gaussian system S_{σ_T} has a simple spectrum. Ageev (2000) shows that Chacon transformation satisfies the SCS property; moreover, in Ageev (2008), he shows that the SCS property is satisfied generically and he gives a construction of a rank one mixing SCS-system

(see also Ryzhikov (2007)). In Lemańczyk and Parreau (2007), it is proved that some special flows considered in section “Special Flows, Flows on Surfaces, and Interval Exchange Transformations” (including the von Neumann class, however, with α having unbounded partial quotients) have the SCS property. It is quite plausible that the SCS property is commonly seen for smooth flows on surfaces.

A classical open problem is whether each ergodic automorphism has a smooth model. While this problem stays open even for so-called dyadic adding machine (ergodic, discrete spectrum automorphism having roots of unity of degree 2^n , $n \geq 1$, as eigenvalues), A. Katok suggested many years ago that one can construct a Kronecker measure so that the corresponding Gaussian system ($\mathbb{Z}$ -action (!)) has a smooth representation on the torus. No “written” proof of this fact yet appeared.

In Vershik (2011), A.M. Vershik sketches a proof of the fact (claimed by himself for decades) that Pascal adic transformation is weakly mixing. Earlier, X. Mela in his PhD showed that this transformation is ergodic and has zero entropy. Moreover, in Janvresse and de la Rue (2004), Janvresse and de la Rue proved the LB property of this map. No complete proof of weak mixing of Pascal adic transformation seems to exist in the published literature.

In Vershik (2005), Vershik proposed to study a new equivalence between measure-preserving automorphisms, called quasi-similarity. Two automorphisms T on $(X, \mathcal{B}, \mu)$ and S on $(Y, \mathcal{C}, \nu)$ are *quasi-similar* if there are Markov operators $J : L^2(X, \mathcal{B}, \mu) \rightarrow L^2(Y, \mathcal{C}, \nu)$ and $K : L^2(Y, \mathcal{C}, \nu) \rightarrow L^2(X, \mathcal{B}, \mu)$, both with dense ranges, intertwining the corresponding Koopman operators U_T and U_S . This new equivalence is strictly stronger than spectral isomorphism but strictly weaker than (weak) isomorphism as shown in Frączek and Lemańczyk (2010) answering one of the questions by Vershik. However, possible invariants for this new equivalence are not well understood. For example, in Frączek and Lemańczyk (2010), it is shown that an automorphism quasi-similar to

a K -automorphism must also be K , but an intriguing question remains whether Bernoulli property is an invariant of quasi-similarity. As weak isomorphism for finite multiplicity automorphisms is in fact isomorphism, another question is whether we can have two non-isomorphic quasi-similar automorphisms with simple spectrum (rank one) (see Problem 4 in Frączek and Lemańczyk (2010)). See also Haase and Moriakov (2018) for the problem of Markov quasi-factors in the class of Abramov automorphisms.

Not too many zero entropy classical dynamical systems with purely Lebesgue spectrum are known: non-zero time automorphisms of horocycle flows and their smooth time changes and even factor of special Gaussian systems. Can we produce a Poissonian suspension over a conservative infinite measure-preserving automorphism having purely Lebesgue spectrum?

Katok and Thouvenot (private communication) considered systems called *infinitely divisible* (ID). These are systems T on $(X, \mathcal{B}, \mu)$ which have a family of factors $\mathcal{B}_\omega$ indexed by $\omega \in \bigcup_{n=0}^{\infty} \{0,1\}^n$ ($\mathcal{B}_\varepsilon = \mathcal{B}$) such that $\mathcal{B}_{\omega 0} \perp \mathcal{B}_{\omega 1}$, $\mathcal{B}_{\omega 0} \vee \mathcal{B}_{\omega 1} = \mathcal{B}_\omega$ and for each $\eta \in \{0,1\}^{\mathbb{N}}$, $\bigcap_{n \in \mathbb{N}} \mathcal{B}_{\eta[0,n]} = \{\emptyset, X\}$. They showed (unpublished) that there are discrete spectrum transformations which are ID and that there are rank one transformations with continuous spectra which are also ID (clearly Gaussian systems are ID). In Lemańczyk et al. (2011), it is proved that dynamical systems coming from stationary ID processes are factors of ID automorphisms; moreover, ID automorphisms are disjoint from all systems having the SCS property. It would be nice to decide whether Koopman representations associated with ID automorphisms satisfy the Kolmogorov group property.

While some lacunary sequences have realization as rigidity sequences for weakly mixing automorphisms, it would be interesting to determine what happens in the non-lacunary case. An especially interesting case is when we consider the non-lacunary multiplicative semigroup $\{2^i 3^j : i, j \geq 0\}$. Is the corresponding sequence a rigidity sequence? What about $\{2^k + 3^\ell : k, \ell \geq 0\}$ (Bergelson et al. 2014).?

Bibliography

- Aaronson J (1979) Rational ergodicity, bounded rational ergodicity, and some continuous measures on the circle. *Isr J Math* 33(3–4):181–197
- Aaronson J (1997) An introduction to infinite ergodic theory. *Mathematical surveys and monographs*, vol 50. American Mathematical Society, Providence
- Aaronson J, Nadkarni MG (1987) L^∞ eigenvalues and L^2 spectra of nonsingular transformations. *Proc Lond Math Soc* 55(3):538–570
- Aaronson J, Weiss B (2000) Remarks on the tightness of cocycles. *Colloq Math* 84(85):363–376
- Aaronson J, Hosseini M, Lemańczyk M (2014) IP-rigidity and eigenvalue groups. *Ergodic Theory Dyn Syst* 34:1057–1076
- Adams T (1998) Smorodinsky's conjecture on rank one systems. *Proc Am Math Soc* 126:739–744
- Adams T (2015) Tower multiplexing and slow weak mixing. *Colloq Math* 138:47–72
- Ageev ON (1988) Dynamical systems with a Lebesgue component of even multiplicity in the spectrum. *Math USSR Sb (N.S.) (Russ)* 136(178):307–319, 430; translation in *Math USSR Sb* 64 (1989), 305–317
- Ageev ON (1999) On ergodic transformations with homogeneous spectrum. *J Dyn Control Syst* 5:149–152
- Ageev ON (2000) On the spectrum of Cartesian powers of classical automorphisms. *Mat Zametki (Russ)* 68:643–647; translation in *Math Notes* 68 (2000):547–551
- Ageev ON (2001) On the multiplicity function of generic group extensions with continuous spectrum. *Ergodic Theory Dyn Syst* 21:321–338
- Ageev ON (2005) The homogeneous spectrum problem in ergodic theory. *Invent Math* 160:417–446
- Ageev ON (2008) Mixing with staircase multiplicity function. *Ergodic Theory Dyn Syst* 28:1687–1700
- Alexeyev VM (1958) Existence of bounded function of the maximal spectral type. *Vestnik Mosc Univ* 5:13–15 and *Ergodic Theory Dyn Syst* 2 (1982):259–261
- Alpern S (1985) Conjecture: in general a mixing transformation is not two-fold mixing. *Ann Probab* 13:310–313
- Anosov DV (2003) On spectral multiplicities in ergodic theory. *Institut im. V.A., Steklova*
- Anzai H (1951) Ergodic skew product transformations on the torus. *Osaka J Math* 3:83–99
- Arni N (2005) Spectral and mixing properties of actions of amenable groups. *Electron Res Announc AMS* 11:57–63
- Arnold VI (1991) Topological and ergodic properties of closed 1-forms with incommensurable periods. *Funktsional Anal Prilozhen (Russ)* 25:1–12
- Avila A, Forni G (2007) Weak mixing for interval exchange transformations and translation flows. *Ann Math* 165:637–664
- Avila A, Forni G, Ulcigrai C (2011) Mixing for the time-changes of Heisenberg nilflows. *J Diff Geom* 89:369–410

- Avila A, Forni G, Ravotti D, Ulcigrai C (2019) Mixing for smooth time-changes of general nilflows, preprint, arXiv:1905.11628
- Badea C, Grivaux S, Matheron E (2018) Rigidity sequences, Kazhdan sets and group topologies on the integers, arXiv:1812.09014
- Balister P, Bollobas B, Morris R, Suhasrabudhe J, Tiba M (2019) Flat Littlewood polynomials exist, arXiv 1907.09464
- Bashtanov AI (2013) Generic mixing transformations are rank 1. *Math Notes* 93:209–216
- Bashtanov AI (2016) Conjugacy classes are dense in the space of mixing $\mathbb{Z}^d$ -actions. *Math Notes* 99:9–23
- Baxter JR (1971) A class of ergodic transformations having simple spectrum. *Proc Am Math Soc* 27:275–279
- Bergelson V, del Junco A, Lemańczyk M, Rosenblatt J (2014) Rigidity and non-recurrence along sequences. *Ergodic Theory Dyn Syst* 34:14641502
- Björklund M, Einsiedler M, Gorodnik A (2020) Quantitative multiple mixing. *J. Eur. Math. Soc. (JEMS)* 22 (5):1475–1529
- Bourgain J (1993) On the spectral type of Ornstein's class of transformations. *Isr J Math* 84:53–63
- Bourgain J (2013) On the correlation of the Moebius function with rank-one systems. *J Anal Math* 120:105–130
- Cambanis S, Podgrski K, Weron A (1995) Chaotic behaviour of infinitely divisible processes. *Stud Math* 115:109–127
- Cecchi PA, Tiedra de Aldecoa R (2016) Furstenberg transformations on Cartesian products of infinite-dimensional tori. *Potent Anal* 44:4351
- Chacon RV (1970) Approximation and spectral multiplicity. In: Dold A, Eckmann B (eds) *Contributions to ergodic theory and probability*. Springer, Berlin, pp 18–27
- Chaika J, Eskin A (2018) Self-joinings for 3-IET's, arXiv:1805.11167v2
- Chaika J, Wright A (2019) A smooth mixing flow on a surface with nondegenerate fixed points. *J Am Math Soc* 32:81–117
- Chaika J, Frączek K, Kanigowski A, Ulcigrai C (2019) Singularity of the spectrum for smooth area-preserving flows in genus two, preprint
- Choe GH (1990) Products of operators with singular continuous spectra. In: *Operator theory: operator algebras and applications, Part 2* (Durham, NH, 1988). *Proceedings of symposia in pure mathematics*, vol 51. American Mathematical Society, Providence, pp 65–68
- Connes A, Woods E (1985) Approximately transitive flows and ITPFI factors. *Ergodic Theory Dyn Syst* 5:203–236
- Cornfeld IP, Fomin SV, Sinai YG (1982) *Ergodic theory*. Springer, New York
- Creutz D, Silva C (2010) Mixing on rank-one transformations. *Stud Math* 199:43–72
- Danilenko AI (2006) Explicit solution of Rokhlin's problem on homogeneous spectrum and applications. *Ergodic Theory Dyn Syst* 26:1467–1490
- Danilenko AI (2010) On new spectral multiplicities for ergodic maps. *Stud Math* 197:57–68
- Danilenko AI (2012) New spectral multiplicities for mixing transformations. *Ergodic Theory Dyn Syst* 32:517–534
- Danilenko AI (2013) A survey on spectral multiplicities of ergodic actions. *Ergodic Theory Dyn Syst* 33:81–117
- Danilenko AI (2014) Mixing actions of the Heisenberg group. *Ergodic Theory Dyn Syst* 34:1142–1167
- Danilenko AI, Lemańczyk M (2005) A class of multipliers for $W^\perp$. *Isr J Math* 148:137–168
- Danilenko AI, Lemańczyk M (2013) Spectral multiplicities for ergodic flows. *Discrete Contin Dyn Syst* 33:4271–4289
- Danilenko AI, Lemańczyk M (2016) Odometer actions of the Heisenberg group. *J Anal Math* 128:107–157
- Danilenko AI, Park KK (2002) Generators and Bernoullian factors for amenable actions and cocycles on their orbits. *Ergodic Theory Dyn Syst* 22:1715–1745
- Danilenko AI, Ryzhikov VV (2010) Spectral multiplicities for infinite measure preserving transformations. *Funct Anal Appl* 44:161–170
- Danilenko AI, Ryzhikov VV (2011) Mixing constructions with infinite invariant measure and spectral multiplicities. *Ergodic Theory Dyn Syst* 31:853–873
- Danilenko AI, Solomko AV (2010) Ergodic Abelian actions with homogenous spectrum. *Contemporary mathematics*, vol 532. American Mathematical Society, Providence, pp 137–148
- de la Rue T (1996) Systèmes dynamiques gaussiens d'entropie nulle, lâchement et non lâchement Bernoulli. *Ergodic Theory Dyn Syst* 16:379–404
- de la Rue T (1998a) Rang des systèmes dynamiques gaussiens. *Isr J Math* 104:261–283
- de la Rue T (1998b) L'ergodicité induit un type spectral maximal équivalent à la mesure de Lebesgue. *Ann Inst Henri Poincaré Probab Statist* 34:249–263
- de la Rue T (1998c) L'induction ne donne pas toutes les mesures spectrales. *Ergodic Theory Dyn Syst* 18:1447–1466
- de la Rue T (2004) An extension which is relatively two-fold mixing but not threefold mixing. *Colloq Math* 101:271–277
- de la Rue T (2009) Joinings in ergodic theory. In: *Encyclopedia of complexity and system science*. Springer, New York
- del Junco A (1976) Transformations with discrete spectrum are stacking transformations. *Can J Math* 28:836–839
- del Junco A (1977) A transformation with simple spectrum which is not of rank one. *Can J Math* 29:655–663
- del Junco A (1981) Disjointness of measure-preserving transformations, minimal self-joinings and category. *Prog Math* 10:81–89
- del Junco A, Lemańczyk M (1992) Generic spectral properties of measure-preserving maps and applications. *Proc Am Math Soc* 115:725–736
- del Junco A, Lemańczyk M (1999) Simple systems are disjoint with Gaussian systems. *Stud Math* 133:249–256
- del Junco A, Lemańczyk M (2005) Joinings of distally simple systems, preprint

- del Junco A, Rudolph D (1987) On ergodic actions whose self-joinings are graphs. *Ergodic Theory Dyn Syst* 7:531–557
- Derriennic Y, Frączek K, Lemańczyk M, Parreau F (2008) Ergodic automorphisms whose weak closure of off-diagonal measures consists of ergodic self-joinings. *Colloq Math* 110:81–115
- Dooley A, Golodets VY (2002) The spectrum of completely positive entropy actions of countable amenable groups. *J Funct Anal* 196:1–18
- Dooley A, Quas A (2005) Approximate transitivity for zero-entropy systems. *Ergodic Theory Dyn Syst* 25:443–453
- Downarowicz T, Kwiatkowski J (2000) Spectral isomorphism of Morse flows. *Fundam Math* 163:193–213
- Downarowicz T, Kwiatkowski J (2002) Weak Closure theorem fails for $\mathbb{Z}^d$ -actions. *Stud Math* 153:115–125
- Eisner T, Grivaux S (2011) Hilbertian Jamison sequences and rigid dynamical systems. *J Funct Anal* 261:2013–2052
- El Abdalaoui H (2000) On the spectrum of the powers of Ornstein transformations. *Sankhya Ser A* 62:291–306. *Ergodic theory and harmonic analysis* (Mumbai, 1999)
- El Abdalaoui H (2007) A new class of rank 1 transformations with singular spectrum. *Ergodic Theory Dyn Syst* 27:1541
- El Abdalaoui H (2015) Ergodic Banach problem on simple Lebesgue spectrum and flat polynomials, arXiv 1508.06439v4
- El Abdalaoui H, Parreau F, Prikhodko A (2006) A new class of Ornstein transformations with singular spectrum. *Ann Inst Henri Poincaré Probab Stat* 42:671–681
- El Abdalaoui H, Lemańczyk M, de la Rue T (2014) On spectral disjointness of powers for rank one transformations and Möbius orthogonality. *J Funct Anal* 266:284–317
- El Abdalaoui H, Kasjan S, Lemańczyk M (2016) 0-1 sequences of the Thue-Morse type and Sarnaks conjecture. *Proc Am Math Soc* 144:161–176
- Fayad B (2001a) Partially mixing and locally rank 1 smooth transformations and flows on the torus T^d , $d \geq 3$. *J Lond Math Soc* 64(2):637–654
- Fayad B (2001b) Polynomial decay of correlations for a class of smooth flows on the two torus. *Bull Soc Math Fr* 129:487–503
- Fayad B (2002a) Skew products over translations on T^d , $d \geq 2$. *Proc Am Math Soc* 130:103–109
- Fayad B (2002b) Analytic mixing reparametrizations of irrational flows. *Ergodic Theory Dyn Syst* 22:437–468
- Fayad B (2005) Rank one and mixing differentiable flows. *Invent Math* 160:305–340
- Fayad B (2006) Smooth mixing flows with purely singular spectra. *Duke Math J* 132:371–391
- Fayad B, Kanigowski A (2015) Rigidity times for a weakly mixing dynamical system which are not rigidity times for any irrational rotation. *Ergodic Theory Dyn Syst* 35:2529–2534
- Fayad B, Kanigowski A (2016) Multiple mixing for a class of conservative surface flows. *Invent Math* 203(2):555–614
- Fayad B, Krikorian R (2018) Some questions around quasi-periodic dynamics, arXiv:1809.10375
- Fayad B, Thouvenot J-P (2014) On the convergence to 0 of $m_n \xi \bmod 1$. *Acta Arith* 165:327332
- Fayad B, Windsor A (2007) A dichotomy between discrete and continuous spectrum for a class of special flows over rotations. *J Modern Dyn* 1:107–122
- Fayad B, Katok AB, Windsor A (2001) Mixed spectrum reparameterizations of linear flows on T^2 , Dedicated to the memory of I. G. Petrovskii on the occasion of his 100th anniversary. *Mosc Math J* 1:521–537
- Fayad B, Forni G, Kanigowski A (2016) Lebesgue spectrum of countable multiplicity for conservative flows on the torus, submitted
- Feldman J (1976) New K-automorphisms and a problem of Kakutani. *Isr J Math* 24:16–38
- Ferenczi S (1984) Systèmes localement de rang un. *Ann Inst Henri Poincaré Probab Statist* 20:35–51
- Ferenczi S (1985) Systèmes de rang un gauche. *Ann Inst Henri Poincaré Probab Statist* 21:177–186
- Ferenczi S (1997) Systems of finite rank. *Colloq Math* 73:35–65
- Ferenczi S, Lemańczyk M (1991) Rank is not a spectral invariant. *Stud Math* 98:227–230
- Ferenczi S, Holton C, Zamboni LQ (2004) Structure of three-interval exchange transformations III: ergodic and spectral properties. *J Anal Math* 93:103–138
- Ferenczi S, Holton C, Zamboni LQ (2005) Joinings of three-interval exchange transformations. *Ergodic Theory Dyn Syst* 25:483–502
- Filipowicz I (1997) Product $\mathbb{Z}^d$ -actions on a Lebesgue space and their applications. *Stud Math* 122:289–298
- Flaminio L, Forni G (2003) Invariant distributions and time averages for horocycle flows. *Duke Math J* 119(3):465–526
- Flaminio L, Forni G (2019) Orthogonal powers and Möbius conjecture for smooth time changes of horocycle flows. *Electron Res Announc Math Sci* 26:16–23
- Foias C, Stratila S (1968) Ensembles de Kronecker dans la théorie ergodique. *C R Acad Sci Paris, Ser A-B* 267: A166–A168
- Forni G, Kanigowski A (2017) Time-changes of Heisenberg nilflows, preprint, arXiv:1711.05543
- Forni G, Kanigowski A (2020) Multiple mixing and disjointness for time changes of bounded-type Heisenberg nilflows. *J Éc polytech Math* 7:63–91
- Forni G, Ulcigrai C (2012) Time-changes of horocycle flows. *J Modern Dyn* 6(2):251–273
- Frączek K (1997) On a function that realizes the maximal spectral type. *Stud Math* 124:1–7
- Frączek K (2000) Circle extensions of $\mathbb{Z}^d$ -rotations on the d -dimensional torus. *J Lond Math Soc* 61(2):139–162
- Frączek K, Lemańczyk M (2003) On symmetric logarithm and some old examples in smooth ergodic theory. *Fundam Math* 180:241–255
- Frączek K, Lemańczyk M (2004) A class of special flows over irrational rotations which is disjoint from mixing flows. *Ergodic Theory Dyn Syst* 24:1083–1095

- Frączek K, Lemańczyk M (2005) On disjointness properties of some smooth flows. *Fundam Math* 185:117–142
- Frączek K, Lemańczyk M (2006) On mild mixing of special flows over irrational rotations under piecewise smooth functions. *Ergodic Theory Dyn Syst* 26:719–738
- Frączek K, Lemańczyk M (2010) A note on quasi-similarity of Koopman operators. *J Lond Math Soc* 82(2):361–375
- Frączek K, Lemańczyk M, Lesigne E (2007) Mild mixing property for special flows under piecewise constant functions. *Discrete Contin Dyn Syst* 19:691–710
- Friedman NA, Ornstein DS (1973) Ergodic transformations induce mixing transformations. *Adv Math* 10:147–163
- Furstenberg H (1967) Disjointness in ergodic theory, minimal sets, and a problem of Diophantine approximation. *Math Syst Theory* 1:1–49
- Furstenberg H (1981) Recurrence in ergodic theory and combinatorial number theory. Princeton University Press, Princeton
- Furstenberg H, Weiss B (1978) The finite multipliers of infinite ergodic transformations. *Lect Notes Math* 668:127–132
- Glaser E (1994) On the class of multipliers for $W^\perp$. *Ergodic Theory Dyn Syst* 14:129–140
- Glaser E (2003) Ergodic theory via joinings. *Mathematical surveys and monographs*, vol 101. AMS, Providence
- Glaser E, Weiss B (1989) Processes disjoint from weak mixing. *Trans Am Math Soc* 316:689–703
- Glimm JG (1960) On a certain class of operator algebras. *Trans Am Math Soc* 95:318–340
- Goodson GR (1999) A survey of recent results in the spectral theory of ergodic dynamical systems. *J Dyn Control Syst* 5:173–226
- Goodson GR, Kwiatkowski J, Lemańczyk M, Liardet P (1992) On the multiplicity function of ergodic group extensions of rotations. *Stud Math* 102:157–174
- Griesmer J (2019) Recurrence, rigidity, and popular differences. *Ergodic Theory Dyn Syst* 39:1299 (to appear)
- Gromov AL (1991) Spectral classification of some types of unitary weighted shift operators. *Algebra Anal (Russ)* 3:62–87; translation in *St. Petersburg Math J* 3 (1992):997–1021
- Guenais M (1998) Une majoration de la multiplicité spectrale d'opérateurs associés à des cocycles réguliers. *Isr J Math* 105:263–284
- Guenais M (1999) Morse cocycles and simple Lebesgue spectrum. *Ergodic Theory Dyn Syst* 19:437–446
- Guenais M, Parreau F (2005) Valeurs propres de transformations liées aux rotations irrationnelles et aux fonctions en escalier, preprint
- Haase M, Moriyakov N (2018) On systems with quasi-discrete spectrum. *Stud Math* 241:173–199
- Hahn F, Parry W (1968) Some characteristic properties of dynamical systems with quasi-discrete spectra. *Math Syst Theory* 2:179–190
- Hasselblatt B, Katok AB (2002) Principal Structures. *Handbook of dynamical systems*, vol 1A. North-Holland, Amsterdam, pp 1–203
- Helson H (1986) Cocycles on the circle. *J Oper Theory* 16:189–199
- Helson H, Parry W (1978) Cocycles and spectra. *Arkiv Math* 16:195–206
- Herman M (1979) Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations. *Publ Math IHES* 49:5–234
- Host B (1991) Mixing of all orders and pairwise independent joinings of systems with singular spectrum. *Isr J Math* 76:289–298
- Host B, Mela F, Parreau F (1991) Non-singular transformations and spectral analysis of measures. *Bull Soc Math Fr* 119:33–90
- Indukaev FK (2007) The twisted Burnside theory for the discrete Heisenberg group and for the wreath products of some groups. *Mosc Univ Math Bull* 62:219–227
- Iwanik A (1991) The problem of L^p -simple spectrum for ergodic group automorphisms. *Bull Soc Math Fr* 119:91–96
- Iwanik A (1992) Positive entropy implies infinite L^p -multiplicity for $p > 1$. *Ergodic theory and related topics, III (Gstrow, 1990)*. *Lecture notes in mathematics*, vol 1514. Springer, Berlin, pp 124–127
- Iwanik A (1997) Anzai skew products with Lebesgue component of infinite multiplicity. *Bull Lond Math Soc* 29:195–199
- Iwanik A, de Sam Lazaro J (1991) Sur la multiplicité L^p d'un automorphisme gaussien. *C R Acad Sci Paris Sér I* 312:875–876
- Iwanik A, Lemańczyk M, Rudolph D (1993) Absolutely continuous cocycles over irrational rotations. *Isr J Math* 83:73–95
- Iwanik A, Lemańczyk M, de Sam Lazaro J, de la Rue T (1997) Quelques remarques sur les facteurs des systèmes dynamiques gaussiens. *Stud Math* 125:247–254
- Iwanik A, Lemańczyk M, Mauduit C (1999) Piecewise absolutely continuous cocycles over irrational rotations. *J Lond Math Soc* 59(2):171–187
- Janicki A, Weron A (1994) Simulation and chaotic behavior of α -stable stochastic processes. *Monographs and textbooks in pure and applied mathematics*, vol 178. Marcel Dekker, New York
- Janvresse E, de la Rue T (2004) The Pascal adic transformation is loosely Bernoulli. *Ann Inst Henri Poincaré Probab Statist* 40:133–139
- Janvresse E, de la Rue T, Ryzhikov VV (2012) Around King's rank one theorems: flows and $\mathbb{Z}^n$ -actions. In: *Dynamical systems and group actions*. *Contemporary mathematics*, vol 567. American Mathematical Society, Providence, pp 143–161
- Janvresse E, Prikhod'ko AA, de la Rue T, Ryzhikov VV (2015) Weak limits of powers of Chacon's automorphism. *Ergodic Theory Dyn Syst* 35:128141

- Janvresse E, Roy E, de la Rue T (2017) Poisson suspensions and Sushis. *Ann Sci École Norm Supérieure* 50(4):1301–1334
- Kachurovskii AG (1990) A property of operators generated by ergodic automorphisms. *Optimizatsiya (Russ)* 47(64):122–125
- Kalikow S (1984) Two fold mixing implies three fold mixing for rank one transformations. *Ergodic Theory Dyn Syst* 4:237–259
- Kamae T. Spectral properties of automata generating sequences, unpublished preprint
- Kamiński B (1981) The theory of invariant partitions for $\mathbb{Z}^d$ -actions. *Bull Acad Polon Sci Ser Sci Math* 29:349–362
- Kanigowski A, de la Rue T (2019) Product of two staircase rank one transformations that is not loosely Bernoulli, accepted in *J. d'Analyse Math.*, arXiv:1812.08027
- Kanigowski A, Kułaga-Przymus J (2016) Ratner's property and mild mixing for smooth flows on surfaces. *Ergodic Theory Dyn Syst* 36:2512–2537
- Kanigowski A, Kułaga-Przymus J, Ulcigrai C (2019) Multiple mixing and parabolic divergence in smooth area-preserving flows on higher genus surfaces. *J Eur Math Soc (JEMS)* 21(12):3797–3855
- Kanigowski A, Lemańczyk M, Ulcigrai C (2018) On disjointness properties of some parabolic flows, preprint, arXiv:1810.11576v1
- Katok AB (1977) Monotone equivalence in ergodic theory. *Izv Akad Nauk SSSR Ser Math (Russ)* 41:104–157
- Katok AB (1980) Interval exchange transformations and some special flows are not mixing. *Isr J Math* 35:301–310
- Katok AB (2001) Cocycles, cohomology and combinatorial constructions in ergodic theory. In: *Collaboration with Robinson EA Jr, Proceedings of symposium on pure mathematics*, vol 69. Smooth ergodic theory and its applications, Seattle, 1999. American Mathematical Society, Providence, pp 107–173
- Katok AB (2003a) Constructions in ergodic theory. Unpublished lecture notes
- Katok AB (2003b) Combinatorial constructions in ergodic theory and dynamics. University lecture series, vol 30. American Mathematical Society, Providence
- Katok A, Lemańczyk M (2009) Some new cases of realization of spectral multiplicity function for ergodic transformations. *Fundam Math* 206:185–215
- Katok AB, Stepin AM (1967) Approximations in ergodic theory. *Uspekhi Mat Nauk (Russ)* 22(137):81–106
- Katok A, Thouvenot J-P (2006) Spectral properties and combinatorial constructions in Ergodic theory. In: *Handbook of dynamical systems*, vol 1B. Elsevier B. V., Amsterdam, pp 649–743
- Katok AB, Zemlyakov AN (1975) Topological transitivity of billiards in polygons. *Mat Zametki (Russ)* 18:291–300
- Keane M (1968) Generalized Morse sequences. *Z Wahr Verw Gebiete* 10:335–353
- Keane M (1975) Interval exchange transformations. *Math Z* 141:25–31
- Khanin KM, Sinai YG (1992) Mixing of some classes of special flows over rotations of the circle. *Funct Anal Appl* 26:155–169
- King J (1986) The commutant is the weak closure of the powers, for rank-1 transformations. *Ergodic Theory Dyn Syst* 6:363–384
- King JL (1988) Joining-rank and the structure of finite rank rank mixing transformations. *J Anal Math* 51:182–227
- King JL (2001) Flat stacks, joining-closure and genericity, preprint
- Kirillov AA (2004) Lectures on the orbit method. Series: Graduate studies in mathematics, vol 64. American Mathematical Society, Providence
- Klimes I (1996) The spectral type of the staircase transformation. *Tohoku Math J* 48:247–248
- Klimes I, Reinhold K (1997) Rank one transformations with singular spectral type. *Isr J Math* 98:1–14
- Koçergin AV (1972) On the absence of mixing in special flows over the rotation of a circle and in flows on a two-dimensional torus. *Dokl Akad Nauk SSSR (Russ)* 205:949–952
- Koçergin AV (1975) Mixing in special flows over rearrangement of segments and in smooth flows on surfaces. *Math-USSR Acad Sci* 25:441–469
- Koçergin AV (1976a) Non-degenerate saddles and absence of mixing. *Mat Zametki (Russ)* 19:453–468
- Koçergin AV (1976b) On the homology of function over dynamical systems. *Dokl Akad Nauk SSSR* 231:795–798
- Koçergin AV (2002) A mixing special flow over a rotation of the circle with an almost Lipschitz function. *Sb Math* 193:359–385
- Koçergin AV (2004) Nondegenerate fixed points and mixing in flows on a two-dimensional torus. II. *Mat Sb (Russ)* 195:15–46
- Koopman BO (1931) Hamiltonian systems and transformations in Hilbert space. *Proc Natl Acad Sci USA* 17:315–318
- Kuipers L, Niederreiter H (1974) Uniform distribution of sequences. Wiley, New York
- Kułaga J, Parreau F (2012) Disjointness properties for Cartesian products of weakly mixing systems. *Colloq Math* 128:153–177
- Kułaga-Przymus J, Lemańczyk M (2019) Sarnak's conjecture from ergodic theory point of view. In: *Encyclopedia of complexity and system science*
- Kushnirenko AG (1974) Spectral properties of some dynamical systems with polynomial divergence of orbits. *Vestn Mosk Univ* 1–3:101–108
- Kwiatkowski J (1981) Spectral isomorphism of Morse dynamical systems. *Bull Acad Polon Sci Sér Sci Math* 29:105–114
- Kwiatkowski J Jr, Lemańczyk M (1995) On the multiplicity function of ergodic group extensions. II. *Stud Math* 116:207–215
- Kwiatkowski J, Lacroix Y (1997) Multiplicity, rank pairs. *J Anal Math* 71:205–235
- Lemańczyk M (1988) Toeplitz Z_2 -extensions. *Ann Inst Henri Poincaré* 24:1–43

- Lemańczyk M (1996) Introduction to ergodic theory from the point of view of the spectral theory. Lecture notes of the tenth KAIST mathematics workshop, Taejon, pp 1–153
- Lemańczyk M (2000) Sur l'absence de mélange pour des flots spéciaux au dessus d'une rotation irrationnelle. *Colloq Math* 84(85):29–41
- Lemańczyk M (2009) Spectral theory of dynamical systems. In: *Encyclopedia of complexity and system science*. Springer, New York, pp 8554–8575
- Lemańczyk M, de Sam Lazaro J (1997) Spectral analysis of certain compact factors for Gaussian dynamical systems. *Isr J Math* 98:307–328
- Lemańczyk M, Lesigne E (2001) Ergodicity of Rokhlin cocycles. *J Anal Math* 85:43–86
- Lemańczyk M, Mauduit C (1994) Ergodicity of a class of cocycles over irrational rotations. *J Lond Math Soc* 49:124–132
- Lemańczyk M, Parreau F (2003) Rokhlin extensions and lifting disjointness. *Ergodic Theory Dyn Syst* 23:1525–1550
- Lemańczyk M, Parreau F (2007) Special flows over irrational rotation with simple convolution property, preprint
- Lemańczyk M, Parreau F (2012) Lifting mixing properties by Rokhlin cocycles. *Ergodic Theory Dyn Syst* 32:763–784
- Lemańczyk M, Sikorski A (1987) A class of not local rank one automorphisms arising from continuous substitutions. *Probab Theory Relat Fields* 76:421–428
- Lemańczyk M, Wasieczko M (2006) A new proof of Alexeyev's theorem, preprint
- Lemańczyk M, Wysokińska M (2007) On analytic flows on the torus which are disjoint from systems of probabilistic origin. *Fundam Math* 195:97–124
- Lemańczyk M, Parreau F, Thouvenot J-P (2000) Gaussian automorphisms whose ergodic self-joinings are Gaussian. *Fundam Math* 164:253–293
- Lemańczyk M, Thouvenot J-P, Weiss B (2002) Relative discrete spectrum and joinings. *Monatsh Math* 137:57–75
- Lemańczyk M, Mentzen MK, Nakada H (2003) Semi-simple extensions of irrational rotations. *Stud Math* 156:31–57
- Lemańczyk M, Parreau F, Roy E (2011) Systems with simple convolutions, distal simplicity and disjointness with infinitely divisible systems. *Proc Am Math Soc* 139:185–199
- Leonov VP (1960) The use of the characteristic functional and semi-invariants in the ergodic theory of stationary processes. *Dokl Akad Nauk SSSR* 133:523–526. *Sov Math* 1 (1960):878–881
- Lightwood S, Sahin A, Ugarcovici I (2014) The structure and the spectrum of Heisenberg odometers. *Proc Am Math Soc* 142:2429–2443
- Mackey GW (1964) Ergodic transformation groups with a pure point spectrum. *Ill J Math* 8:593–600
- Marcus B (1977) Ergodic properties of horocycle flows for surfaces of negative curvature. *Ann Math* 105(2):81–105
- Maruyama G (1970) Infinitely divisible processes. *Theory Probab Appl* 15(1):1–22
- Masur H (1982) Interval exchange transformations and measured foliations. *Ann Math* 115:169–200
- Mathew J, Nadkarni MG (1984) Measure-preserving transformation whose spectrum has Lebesgue component of multiplicity two. *Bull Lond Math Soc* 16:402–406
- Medina H (1994) Spectral types of unitary operators arising from irrational rotations on the circle group. *Mich Math J* 41:39–49
- Mentzen MK (1988) Some examples of automorphisms with rank r and simple spectrum. *Bull Pol Acad Sci* 7–8:417–424
- Nadkarni MG (1998) Spectral theory of dynamical systems. Hindustan Book Agency, New Delhi
- Newton D (1966) On Gaussian processes with simple spectrum. *Z Wahr Verw Gebiete* 5:207–209
- Newton D, Parry W (1966) On a factor automorphism of a normal dynamical system. *Ann Math Statist* 37:1528–1533
- Ornstein D (1970) On the root problem in ergodic theory. In: *Proceedings of 6th Berkeley symposium on mathematical statistics and probability*. University California Press, Berkeley, pp 348–356
- Ornstein D, Weiss B (1987) Entropy and isomorphism theorems for actions of amenable groups. *J Anal Math* 48:1–141
- Ornstein D, Rudolph D, Weiss B (1982) Equivalence of measure preserving transformations. *Mem Am Math Soc* 37(262):1–116
- Parreau F (2000) On the Foiaş and Stratila theorem. In: *Proceedings of the conference on ergodic theory*, Toruń
- Parreau F, Roy E (2015) Prime Poisson suspensions. *Ergodic Theory Dyn Syst* 35:2216–2230
- Parry W (1970) Spectral analysis of G -extensions of dynamical systems. *Topology* 9:217–224
- Parry W (1981) Topics in ergodic theory. Cambridge tracts in mathematics, vol 75. Cambridge University Press, Cambridge/New York
- Petersen K (1983) Ergodic theory. Cambridge University Press, Cambridge
- Prikhodko AA (2013) Littlewood polynomials and their applications to the spectral theory of dynamical systems. *Mat Sb* 204:135–160, translation in *Sb Math* 204 (2013)
- Prikhodko AA, Ryzhikov VV (2000) Disjointness of the convolutions for Chacon's automorphism. Dedicated to the memory of Anzelm Iwanik. *Colloq Math* 84(85):67–74
- Quefelec M (1988) Substitution dynamical systems – spectral analysis. Lecture notes in mathematics, vol 1294. Springer, Berlin
- Ratner M (1978) Horocycle flows are loosely Bernoulli. *Isr J Math* 31:122–132
- Ratner M (1979) The Cartesian square of the horocycle flow is not loosely Bernoulli. *Isr J Math* 34(1–2):72–96
- Ratner M (1983) Horocycle flows, joinings and rigidity of products. *Ann Math* 118:277–313

- Ratner M (1986) Rigidity of time changes for horocycle flows. *Acta Math* 156:1–32
- Ratner M (1987) Rigid reparametrizations and cohomology for horocycle flows. *Invent Math* 88(2):341–374
- Rauzy G (1979) Echanges d'intervalles et transformations induites. *Acta Arith* 34:315–328
- Ravotti D (2017) Quantitative mixing for locally Hamiltonian flows with saddle loops on compact surfaces. *Ann Henri Poincaré* 18(12):3815–3861
- Ravotti D (2019) Mixing for suspension flows over skew-translations and time-changes of quasi-abelian filiform nilflows. *Ergodic Theory Dyn Syst*, published online at <https://doi.org/10.1017/etds.2018.19>
- Robinson EA Jr (1983) Ergodic measure preserving transformations with arbitrary finite spectral multiplicities. *Invent Math* 72:299–314
- Robinson EA Jr (1986) Transformations with highly non-homogeneous spectrum of finite multiplicity. *Isr J Math* 56:75–88
- Robinson EA Jr (1988) Nonabelian extensions have non-simple spectrum. *Compos Math* 65:155–170
- Robinson EA Jr (1992) A general condition for lifting theorems. *Trans Am Math Soc* 330:725–755
- Rosiński J, Żak T (1996) Simple condition for mixing of infinitely divisible processes. *Stoch Process Appl* 61:277–288
- Rosinski J, Żak T (1997) The equivalence of ergodicity and weak mixing for infinitely divisible processes. *J Theor Probab* 10:73–86
- Roy E (2005) Mesures de Poisson, infinie divisibilit et propriés ergodiques. Thèse de doctorat de l'Université Paris 6
- Roy E (2007) Ergodic properties of Poissonian ID processes. *Ann Probab* 35:551–576
- Royden HL (1968) Real analysis. MacMillan, New York
- Rudin W (1962) Fourier analysis on groups. In: *Interscience tracts in pure and applied mathematics*, vol 12. Interscience Publishers (a division of John Wiley and Sons), New York/London
- Rudolph DJ (1979) An example of a measure-preserving map with minimal self-joinings and applications. *J Anal Math* 35:97–122
- Rudolph D (1985) k -fold mixing lifts to weakly mixing isometric extensions. *Ergodic Theory Dyn Syst* 5:445–447
- Rudolph D (1986) $\mathbb{Z}^n$ and $\mathbb{R}^n$ cocycle extension and complementary algebras. *Ergodic Theory Dyn Syst* 6:583–599
- Rudolph D (1990) Fundamentals of measurable dynamics. Oxford Science Publications, Oxford
- Rudolph D (2004) Pointwise and L^1 mixing relative to a sub-sigma algebra. *Ill J Math* 48:505–517
- Rudolph D, Weiss B (2000) Entropy and mixing for amenable group actions. *Ann Math* 151(2):1119–1150
- Ryzhikov VV (1991) Joinings of dynamical systems. Approximations and mixing. *Uspekhi Mat Nauk (Russ)* 46(5(281)):177–178
- Ryzhikov VV (1992) Mixing, rank and minimal self-joining of actions with invariant measure. *Mat Sb (Russ)* 183:133–160
- Ryzhikov VV (1994a) The absence of mixing in special flows over rearrangements of segments. *Mat Zametki (Russ)* 55:146–149
- Ryzhikov VV (1994b) Skew products and multiple mixing of dynamical systems. *Russ Math Surv* 49:170–171
- Ryzhikov VV (1996) Stochastic intertwining and multiple mixing of dynamical systems. *J Dyn Control Syst* 2:1–19
- Ryzhikov VV (1999) Transformations having homogeneous spectra. *J Dyn Control Syst* 5:145–148
- Ryzhikov VV (2000) The Rokhlin problem on multiple mixing in the class of actions of positive local rank. *Funktsional Anal Prilozhen (Russ)* 34:90–93
- Ryzhikov VV (2007) Weak limits of powers, simple spectrum symmetric products and mixing rank one constructions. *Math Sb* 198:733–754
- Ryzhikov VV (2009) Spectral multiplicities and asymptotic operator properties of actions with invariant measure. *Sb Math* 200:1833–1845
- Ryzhikov VV (2014) On spectral multiplicities of Gaussian actions, arXiv 1406.3321
- Ryzhikov VV, Thouvenot J-P (2006) Disjointness, divisibility, and quasi-simplicity of measure-preserving actions. *Funktsional Anal Prilozhen (Russ)* 40:85–89
- Ryzhikov VV, Troitskaya AE (2016) Mixing flows with a homogeneous spectrum of multiplicity 2. *Fundam Prikl Mat (Russ)* 21:191–197
- Sato K-I (1999) Lévy processes and infinitely divisible distributions. Cambridge University Press, Cambridge
- Schmidt K (1977) Cocycles of ergodic transformation groups. *Lecture notes in math*, vol 1. Mac Millan of India, New Delhi
- Schmidt K (2002) Dispersing cocycles and mixing flows under functions. *Fundam Math* 173:191–199
- Schmidt K, Walters P (1982) Mildly mixing actions of locally compact groups. *Proc Lond Math Soc* 45(3):506–518
- Shklover M (1967) Classical dynamical systems on the torus with continuous spectrum. *Izv Vys Uchebn Zaved Mat (Russ)* 10(65):113–124
- Simonelli LD (2018) Absolutely continuous spectrum for parabolic flows/maps. *Discrete Contin Dyn Syst* 38(1):263–292
- Sinai YG (1994) Topics in ergodic theory. Princeton University Press, Princeton
- Smorodinsky M, Thouvenot J-P (1979) Bernoulli factors that span a transformation. *Isr J Math* 32:39–43
- Solomko AV (2012) New spectral multiplicities for ergodic actions. *Stud Math* 208:229–247
- Stepin AM (1986) Spectral properties of generic dynamical systems. *Izv Akad Nauk SSSR Ser Math (Russ)* 50:801–834
- Thoma E (1964) Über unitäre Darstellungen abzählbarer, diskreter Gruppen. *Math Ann* 153:111–138
- Thouvenot J-P (1995) Some properties and applications of joinings in ergodic theory. In: *Ergodic theory and its connections with harmonic analysis*. London Mathematical Society. Cambridge University Press, Cambridge, pp 207–235

- Thouvenot J-P (2000) Les systèmes simples sont disjoints de ceux qui sont infiniment divisibles et plongeables dans un flot. *Colloq Math* 84(85):481–483
- Tiedra de Aldecoa R (2012) Spectral analysis of time-changes of the horocycle flow. *J Modern Dyn* 6(2):275–285
- Tiedra de Aldecoa R (2015a) The absolute continuous spectrum of skew products of compact lie groups. *Isr J Math* 208:323–350
- Tiedra de Aldecoa R (2015b) Commutator methods for the spectral analysis of uniquely ergodic dynamical systems. *Ergodic Theory Dyn Syst* 35:944–967
- Tiedra de Aldecoa R (2017) Commutator criteria for strong mixing. *Ergodic Theory Dyn Syst* 37:308–323
- Tikhonov SV (2002) On a relation between the metric and spectral properties of $\mathbb{Z}^d$ -actions. *Fundam Prikl Mat* 8:1179–1192
- Tikhonov SV (2011) Mixing transformations with homogeneous spectrum. *Mat Sb* 202(8):139–160; English transl in *Sb Math* 202(8) (2011):1231–1252
- Tikhonov SV (2012) Genericity of a multiple mixing. *Uspekhi Mat Nauk* (Russ) 67(4(406)):187–188; translation in *Russian Math Surv* 67(4) (2012): 779–780
- Tikhonov SV (2013) Complete metric on mixing actions of general groups. *J Dyn Control Syst* 19:17–31
- Todd JA (1950) On a conjecture in group theory. *J Lond Math Soc* 25:246
- Ulcigrai C (2007) Mixing of asymmetric logarithmic suspension flows over interval exchange transformations. *Ergodic Theory Dyn Syst* 27:991–1035
- Ulcigrai C (2011) Absence of mixing in area-preserving flows on surfaces. *Ann Math* 173(3):1743–1778
- Veech W (1978) Interval exchange transformations. *J Anal Math* 33:222–272
- Veech WA (1982a) A criterion for a process to be prime. *Monatsh Math* 94:335–341
- Veech W (1982b) Gauss measures for transformations on the space of interval exchange maps. *Ann Math* 115(2):201–242
- Veech W (1984) The metric theory of interval exchange transformations. I. Generic spectral properties. *Am J Math* 106:1331–1359
- Vershik AM (1962a) On the theory of normal dynamic systems. *Math Sov Dokl* 144:625–628
- Vershik AM (1962b) Spectral and metric isomorphism of some normal dynamical systems. *Math Sov Dokl* 144:693–696
- Vershik AM (2005) Polymorphisms, Markov processes, and quasi-similarity. *Discrete Contin Dyn Syst* 13:1305–1324
- Vershik AM (2011) The Pascal automorphism has a continuous spectrum. *Funktsional Anal Prilozhen* (Russ) 45:16–33; translation in *Funct Anal Appl* 45 (2011):173–186
- von Neumann J (1932) Zur Operatorenmethode in der Klassischen Mechanik. *Ann Math* 33:587–642
- Walters P (1982) An introduction to ergodic theory. Springer, New York
- Wysokinska M (2004) A class of real cocycles over an irrational rotation for which Rokhlin cocycle extensions have Lebesgue component in the spectrum. *Topol Methods Nonlinear Anal* 24:387–407
- Wysokińska M (2007) Ergodic properties of skew products and analytic special flows over rotations. PhD thesis, Toruń
- Yassawi R (2003) Multiple mixing and local rank group actions. *Ergodic Theory Dyn Syst* 23:1275–1304
- Zeitz P (1993) The centralizer of a rank one flow. *Isr J Math* 84:129–145

Joinings in Ergodic Theory

Thierry de la Rue

Laboratoire de Mathématiques Raphaël Salem,
CNRS – Université de Rouen, Saint Étienne du
Rouvray, France

Article Outline

Glossary

Definition of the Subject

Introduction

Joinings of Two or More Dynamical Systems

Self-Joinings

Some Applications and Future Directions

Bibliography

Keywords

Measure · Preserving dynamical system ·
Joining · Disjointness · Factor · Minimal self-
joining

Glossary

Disjoint Measure-Preserving Systems

The two measure-preserving dynamical systems $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ are said to be *disjoint* if their only joining is the product measure $\mu \otimes \nu$.

Joining

Let I be a finite or countable set, and for each $i \in I$, let $(X_i, \mathcal{A}_i, \mu_i, T_i)$ be a measure-preserving dynamical system. A *joining* of these systems is a probability measure on the Cartesian product $\prod_{i \in I} X_i$, which has the μ_i s as marginals and which is invariant under the product transformation $\otimes_{i \in I} T_i$.

Marginal of a Probability Measure on a Product Space

Let λ be a probability measure on the Cartesian product of a finite or countable collection of measurable spaces $(\prod_{i \in I} X_i, \otimes_{i \in I} \mathcal{A}_i)$, and let $J = j_1, \dots, j_k$ be a finite subset of I . The *k-fold marginal* of λ on $X_{j_1}, \dots, X_{j_k}$ is the probability measure μ defined by:

$$\forall A_1 \in \mathcal{A}_{j_1}, \dots, A_k \in \mathcal{A}_{j_k},$$

$$\mu(A_1 \times \dots \times A_k) := \lambda \left(A_1 \times \dots \times A_k \times \prod_{i \in I \setminus J} X_i \right).$$

Markov Intertwining

Let $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ be two measure-preserving dynamical systems. We call *Markov intertwining* of T and S any operator $P : L^2(X, \mu) \rightarrow L^2(Y, \nu)$ enjoying the following properties:

- $PU_T = U_S P$, where U_T and U_S are the unitary operators on $L^2(X, \mu)$ and $L^2(Y, \nu)$ associated, respectively, with T and S (i.e., $U_T f(x) = f(Tx)$, and $U_S g(y) = g(Sy)$).

$$P1_X = 1_Y,$$

- $f \geq 0$ implies $Pf \geq 0$, and $g \geq 0$ implies $P^* g \geq 0$, where P^* is the adjoint operator of P .

Minimal Self-Joinings

Let $k \geq 2$ be an integer. The ergodic measure-preserving dynamical system T has *k-fold minimal self-joinings* if, for any ergodic joining λ of k copies of T , we can partition the set $\{1, \dots, k\}$ of coordinates into subsets $J_1, \dots, J_\ell$ such that:

1. For j_1 and j_2 belonging to the same J_i , the marginal of λ on the coordinates j_1 and j_2 is supported on the graph of T^n for some integer n (depending on j_1 and j_2).
2. For $j_1 \in J_1, \dots, j_\ell \in J_\ell$, the coordinates $j_1, \dots, j_\ell$ are independent.

We say that T has *minimal self-joinings* if T has k -fold minimal self-joinings for every $k \geq 2$.

Off-Diagonal Self-Joinings

Let $(X, \mathcal{A}, \mu, T)$ be a measure-preserving dynamical system and S be an invertible measure-preserving transformation of $(X, \mathcal{A}, \mu)$ commuting with T . Then the probability measure Δ_S defined on $X \times X$ by

$$\Delta_S(A \times B) := \mu(A \cap S^{-1}B) \quad (1)$$

is a twofold self-joining of T supported on the graph of S . We call it an *off-diagonal self-joining* of T .

Process in a Measure-Preserving Dynamical Systems

Let $(X, \mathcal{A}, \mu, T)$ be a measure-preserving dynamical system, and let $(E, \mathcal{B}(E))$ be a measurable space (which may be a finite or countable set, or $\mathbb{R}^d$, or $\mathbb{C}^d \dots$). For any E -valued random variable ξ defined on the probability space $(X, \mathcal{A}, \mu)$, we can consider the stochastic process $(\xi_i)_{i \in \mathbb{Z}}$ defined by

$$\xi_i := \xi \circ T^i.$$

Since T preserves the probability measure μ , $(\xi_i)_{i \in \mathbb{Z}}$ is a *stationary process*: For any l and n , the distribution of $(\xi_0, \dots, \xi_l)$ is the same as the probability distribution of $(\xi_n, \dots, \xi_{n+l})$.

Self-Joining

Let T be a measure-preserving dynamical system. A *self-joining* of T is a joining of a family $(X_i, \mathcal{A}_i, \mu_i, T_i)_{i \in I}$ of systems where each T_i is a copy of T . If I is finite and has cardinal k , we speak of a k -fold self-joining of T .

Simplicity

For $k \geq 2$, we say that the ergodic measure-preserving dynamical system T is *k-fold simple* if, for any ergodic joining λ of k copies of T , we can partition the set $\{1, \dots, k\}$ of coordinates into subsets $J_1, \dots, J_\ell$ such that:

1. For j_1 and j_2 belonging to the same J_i , the marginal of λ on the coordinates j_1 and j_2 is

supported on the graph of some $S \in C(T)$ (depending on j_1 and j_2).

2. For $j_1 \in J_1, \dots, j_\ell \in J_\ell$, the coordinates $j_1, \dots, j_\ell$ are independent.

We say that T is *simple* if T is k -fold simple for every $k \geq 2$.

Definition of the Subject

The word *joining* can be considered as the counterpart in ergodic theory of the notion of *coupling* in probability theory (see, e.g., Thorisson 2000): Given two or more processes defined on different spaces, what are the possibilities of embedding them together in the same space? There always exists the solution of making them independent of each other, but interesting cases arise when we can do this in other ways. The notion of joining originates in ergodic theory from pioneering works of Furstenberg (1967), who introduced the fundamental notion of *disjointness*, and Rudolph, who laid the basis of joining theory in his article on minimal self-joinings (Rudolph 1979). It has today become an essential tool in the classification of measure-preserving dynamical systems and in the study of their intrinsic properties.

Introduction

A central question in ergodic theory is to tell when two measure-preserving dynamical systems are essentially the same, i.e., when they are *isomorphic*. When this is not the case, a finer analysis consists in asking what these two systems could share in common: For example, do there exist stationary processes which can be observed in both systems? This latter question can also be asked in the following equivalent way: Do these two systems have a *common factor*? The arithmetical flavor of this question is not fortuitous: There are deep analogies between the arithmetic of integers and the classification of measure-preserving dynamical systems, and these analogies were at the starting point of the study of joinings in ergodic theory.

In the seminal paper (Furstenberg 1967) which introduced the concept of joinings in ergodic theory, Furstenberg observed that two operations can be done with dynamical systems: We can consider the *product* of two dynamical systems, and we can also take a *factor* of a given system. Like the multiplication of integers, the product of dynamical systems is commutative and associative, it possesses a neutral element (the trivial single-point system), and the systems S and T are both factors of their product $S \times T$. It was then natural to introduce the property for two measure-preserving systems to be *relatively prime*. As far as integers are concerned, there are two equivalent ways of characterizing the relative primeness: First, the integers a and b are *relatively prime* if their unique positive common factor is 1. Second, a and b are *relatively prime* if, each time both a and b are factors of an integer c , their product ab is also a factor of c . It is a well-known theorem in number theory that these two properties are equivalent, but this was not clear for their analog in ergodic theory. Furstenberg reckoned that the second way of defining relative primeness was the most interesting property in ergodic theory and called it *disjointness* of measure-preserving systems (we will discuss precisely in section “[From Disjointness to Isomorphism](#)” what the correct analog is in the setting of ergodic theory). He also asked whether the nonexistence of a nontrivial common factor between two systems was equivalent to their disjointness. He was able to prove that disjointness implies the impossibility of a nontrivial common factor, but not the converse. And in fact, the converse turns out to be false: In 1979, Rudolph exhibited a counterexample in his paper introducing the important notion of *minimal self-joinings*. The relationships between disjointness and the lack of common factor will be presented in details in section “[Joinings and Factors](#).”

Given two measure-preserving dynamical systems S and T , the study of their disjointness naturally leads one to consider all the possible ways these two systems can be both seen as factors of a third system. As we shall see, this is precisely the study of their *joinings*. The concept of joining

turns out to be related with many important questions in ergodic theory, and a large number of deep results can be stated and proved inside the theory of joinings. For example, the fact that the dynamical systems S and T are isomorphic is equivalent to the existence of a special joining between S and T , and this can be used to give a joining proof of Krieger’s finite generator theorem, as well as Ornstein’s isomorphism theorem (see section “[Joinings Proofs of Ornstein’s and Krieger’s Theorems](#)”). As it already appears in Furstenberg’s article, joinings provide a powerful tool in the classification of measure-preserving dynamical systems: Many classes of systems can be characterized in terms of their disjointness with other systems. Joinings are also strongly connected with difficult questions arising in the study of almost everywhere convergence of non-conventional averages (see section “[Joinings and Multiple Ergodic Averages](#)”).

Amazingly, a situation in which the study of joinings leads to most interesting results consists in considering two or more *identical* systems. We then speak of the *self-joinings* of the dynamical system T . Again, the study of self-joinings is closely related to many ergodic properties of the system: its mixing properties, the structure of its factors, the transformations which commute with T , and so on. We already mentioned minimal self-joinings, and we will see in section “[Minimal Self-Joinings](#)” how this property may be used to get many interesting examples, such as a transformation with no root, or a process with no nontrivial factor. In the same section, we will also discuss a very interesting generalization of minimal self-joinings: the property of being *simple*.

The range of applications of joinings in ergodic theory is very large; only some of them will be given in section “[Some Applications and Future Directions](#)”: The use of joinings in proving Krieger’s and Ornstein’s theorems, the links between joinings and some questions of pointwise convergence, and the strong connections between the study of self-joinings and Rohlin’s famous question on multifold mixing, which has been opened since 1949 (Rohlin 1949).

In parallel with the concept of joinings of measure-preserving systems, Furstenberg also introduced in Furstenberg (1967) the notion of *topological joinings*, concerning topological dynamical systems (i.e., systems given by a continuous transformation of a compact metric space). In the present article, we have restricted ourselves to the measure-preserving setting. In addition to Furstenberg's paper, applications of topological joinings can be found, for example, in Furstenberg et al. (1973), King (1990), and Weiss (1998).

Joinings of Two or More Dynamical Systems

In the following, we are given a finite or countable family $(X_i, \mathcal{A}_i, \mu_i, T_i)_{i \in I}$ of measure-preserving dynamical systems: T_i is an invertible measure-preserving transformation of the standard Borel probability space $(X_i, \mathcal{A}_i, \mu_i)$. When it is not ambiguous, we shall often use the symbol T_i to denote both the transformation and the system.

A *joining* λ of the T_i s (see the definition in the "Glossary") defines a new measure-preserving dynamical system: The product transformation

$$\bigotimes_{i \in I} T_i : (x_i)_{i \in I} \mapsto (T_i x_i)_{i \in I}$$

acting on the Cartesian product $\prod_{i \in I} X_i$ and preserving the probability measure λ . We will denote this big system by $(\bigotimes_{i \in I} T_i)_\lambda$. Since all marginals of λ are given by the original probabilities μ_i , observing only the coordinate i in the big system is the same as observing only the system T_i . Thus, each system T_i is a factor of $(\bigotimes_{i \in I} T_i)_\lambda$, via the homomorphism π_i which maps any point in the Cartesian product to its i -th coordinate.

Conversely, if we are given a measure-preserving dynamical system $(Z, \mathcal{C}, \rho, R)$ admitting each T_i as a factor via some homomorphism $\varphi_i: Z \rightarrow X_i$, then we can construct the map $\varphi: Z \rightarrow \prod_{i \in I} X_i$ sending z to $(\varphi_i(z))_{i \in I}$. We can easily check that the image of the probability measure ρ is then a joining of the T_i s.

Therefore, studying the joinings of a family of measure-preserving dynamical system amounts to study all the possible ways these systems can be together seen as factors in another big system.

The Set of Joinings

The set of all joinings of the T_i s will be denoted by $J(T_i, i \in I)$. Before anything else, we have to observe that this set is never empty. Indeed, whatever the systems are, the product measure $\bigotimes_{i \in I} \mu_i$ always belongs to this set. Note also that any convex combination of joinings is a joining: $J(T_i, i \in I)$ is a convex set.

The set of joinings is turned into a compact metrizable space, equipped with the topology defined by the following notion of convergence: $\lambda_n \xrightarrow[n \rightarrow \infty]{} \lambda$ if and only if, for all family of measurable subsets $(A_i)_{i \in I} \in \prod_{i \in I} \mathcal{A}_i$, finitely many of them being different from X_i , we have

$$\lambda_n \left(\prod_{i \in I} A_i \right) \xrightarrow[n \rightarrow \infty]{} \lambda \left(\prod_{i \in I} A_i \right). \quad (2)$$

We can easily construct a distance defining this topology by observing that it is enough to check (2) when each of the A_i s is chosen in some countable algebra $\mathcal{C}_i$ generating the σ -algebra $\mathcal{A}_i$. We can also point out that, when the X_i s are themselves compact metric spaces, this topology on the set of joinings is nothing but the restriction to $J(T_i, i \in I)$ of the usual weak* topology.

It is particularly interesting to study *ergodic* joinings of the T_i s, whose set will be denoted by $J_e(T_i, i \in I)$. Since any factor of an ergodic system is itself ergodic, a necessary condition for $J_e(T_i, i \in I)$ not to be empty is that all the T_i s be themselves ergodic. Conversely, if all the T_i s are ergodic, we can prove by considering the ergodic decomposition of the product measure $\bigotimes_{i \in I} \mu_i$ that ergodic joinings do exist: Any ergodic measure appearing in the ergodic decomposition of some joining has to be itself a joining. This result can also be stated in the following way:

Proposition 1

If all the T_i s are ergodic, the set of their ergodic joinings is the set of extremal points in the compact convex set $J(T_i, i \in I)$.

From Disjointness to Isomorphy

In this section, as in many others in this entry, we are focusing on the case where our family of dynamical systems is reduced to two of them. We will then rather call them S and T , standing for $(Y, \mathcal{B}, \nu, S)$ and $(X, \mathcal{A}, \mu, T)$. We are interested here in two extremal cases for the set of joinings $J(T, S)$. The first one occurs when the two systems are as far as possible from each other: They have nothing to share in common, and therefore their set of joinings is reduced to the singleton $\{\mu \otimes \nu\}$: This is called the *disjointness* of S and T . The second one arises when the two systems are isomorphic, and we will see how this property shows through $J(T, S)$.

Disjointness

Many situations where disjointness arises were already given by Furstenberg in (1967). Particularly interesting is the fact that classes of dynamical systems can be characterized through disjointness properties. We list here some of the main examples of disjoint classes of measure-preserving systems.

Theorem 1

1. *T is ergodic if and only if it is disjoint from every identity map.*
2. *T is weakly mixing if and only if it is disjoint from any rotation on the circle.*
3. *T has zero entropy if and only if it is disjoint from any Bernoulli shift.*
4. *T is a K -system if and only if it is disjoint from any zero-entropy system.*

The first result is the easiest, but is quite important, in particular when it is stated in the following form: If λ is a joining of T and S , with T ergodic, and if λ is invariant by $T \times \text{Id}$, then $\lambda = \mu \otimes \nu$.

The second, third, and fourth results were originally proved by Furstenberg. They can also be seen as corollaries of the theorems presented in section “[Joinings and Factors](#),” linking the non-

disjointness property with the existence of a particular factor.

Both the first and the second results can be derived from the next theorem, giving a general spectral condition in which disjointness arises. The proof of this theorem can be found in Thouvenot (1995). It is a direct consequence of the fact that, if f and g are square-integrable functions in a given dynamical system, and if their spectral measures are mutually singular, then f and g are orthogonal in L^2 .

Theorem 2

If the reduced maximum spectral types of T and S are mutually singular, then T and S are disjoint.

As we already said in the introduction, disjointness was recognized by Furstenberg as the most pertinent way to define the analog of the arithmetic property “ a and b are relatively prime” in the context of measure-preserving dynamical systems. We must however point out that the statement:

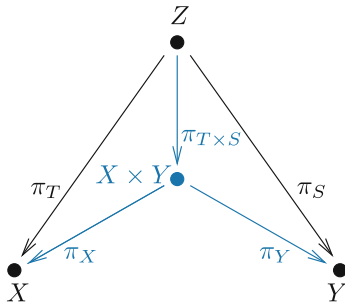
(i) S and T are disjoint is, in general, strictly stronger than the straightforward translation of the arithmetic property.

(ii) Each time both S and T appear as factors in a third dynamical system, then their product $S \times T$ also appears as a factor in this system.

Indeed, contrary to the situation in ordinary arithmetic, there exist nontrivial dynamical systems T which are isomorphic to $T \times T$: For example, this is the case when T is the product of countably many copies of a single nontrivial system. Now, if T is such a system and if we take $S = T$, then S and T do not satisfy statement (i): A nontrivial system is never disjoint from itself, as we will see in the next section. However, they obviously satisfy the statement (ii).

A correct translation of the arithmetic property is the following: S and T are disjoint if and only if, each time T and S appear as factors in some dynamical system through the respective homomorphisms π_T and π_S , $T \times S$ also appears as a factor through a homomorphism $\pi_{T \times S}$ such that $\pi_X \circ \pi_{T \times S} = \pi_T$ and $\pi_Y \circ \pi_{T \times S} = \pi_S$, where π_X and π_Y are the projections on the coordinates

in the Cartesian product $X \times Y$ (see the diagram below).



Joinings and Isomorphism

We first introduce some notations: For any probability measure λ on a measurable space, let ' $A \stackrel{\lambda}{=} B$ ' stand for ' $\lambda(A \Delta B) = 0$ '. Similarly, if $\mathcal{C}$ and $\mathcal{D}$ are σ -algebras of measurable sets, we write ' $\mathcal{C} \stackrel{\lambda}{\subset} \mathcal{D}$ ' if, for any $C \in \mathcal{C}$, we can find some $D \in \mathcal{D}$ such that $C \stackrel{\lambda}{=} D$, and by ' $\mathcal{C} \stackrel{\lambda}{\subset} \mathcal{D}$ ' we naturally mean that both $\mathcal{C} \stackrel{\lambda}{\subset} \mathcal{D}$ and $\mathcal{D} \stackrel{\lambda}{\subset} \mathcal{C}$ hold.

Let us assume now that our two systems S and T are isomorphic: This means that we can find some measurable one-to-one map $\varphi : X \rightarrow Y$, with $T(\mu) = \nu$, and $\varphi \circ T = S \circ \varphi$. With such a φ , we construct the measurable map $\psi : X \rightarrow X \times Y$ by setting

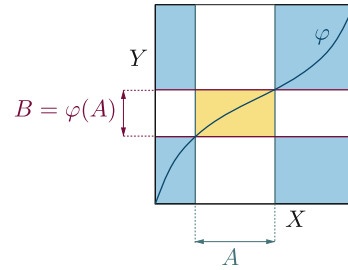
$$\psi(x) := ((x, \varphi(x))).$$

Let Δ_φ be the image measure of μ by ψ . This measure is supported on the graph of φ and is also characterized by

$$\begin{aligned} \forall A \in \mathcal{A}, \forall B \in \mathcal{B}, \Delta_\varphi(A \times B) \\ = \mu(A \cap \varphi^{-1}(B)). \end{aligned} \quad (3)$$

We can easily check that, φ being an isomorphism of T and S , Δ_φ is a joining of T and S . And this joining satisfies very special properties (Fig. 1):

- For any measurable $A \subset X$, $A \times Y \stackrel{\Delta_\varphi}{=} X \times \varphi(A)$.
- Conversely, for any measurable $B \subset Y$, $X \times B \stackrel{\Delta_\varphi}{=} \varphi^{-1}(B) \times Y$.



Joinings in Ergodic Theory, Fig. 1 The joining Δ_φ identifies the sets $A \times Y$ and $X \times \varphi(A)$

Thus, in the case where S and T are isomorphic, we can find a special joining of S and T , which is supported on the graph of an isomorphism and which identifies the two σ -algebras generated by the two coordinates. What is remarkable is that the converse is true: The existence of an isomorphism between S and T is characterized by the existence of such a joining, and we have the following theorem:

Theorem 3

The measure-preserving dynamical systems S and T are isomorphic if and only if there exists a joining λ of S and T such that

$$\{X, \emptyset\} \otimes \mathcal{B} \stackrel{\lambda}{=} \mathcal{A} \otimes \{Y, \emptyset\}. \quad (4)$$

When λ is such a joining, it is supported on the graph of an isomorphism of T and S , and both systems are isomorphic to the joint system $(T \otimes S)_\lambda$.

This theorem finds nice applications in the proof of classical isomorphism results. For example, it can be used to prove that two discrete-spectrum systems which are spectrally isomorphic are isomorphic (see Thouvenot 1995 or de la Rue 2006b). We will also see in section “[Joinings Proofs of Ornstein’s and Krieger’s Theorems](#)” how it can be applied in the proofs of Krieger’s and Ornstein’s deep theorems.

Consider now the case where T and S are no longer isomorphic, but where S is only a factor of T . Then we have a factor map $\pi : X \rightarrow Y$ which has the same properties as an isomorphism φ , except that it is not one-to-one (π is only onto). The measure Δ_π , constructed in the same way as Δ_φ ,

is still a joining supported on the graph of π , but it does not identify the two σ -algebras generated by the two coordinates any more: Instead of Condition (4), Δ_π only satisfies the weaker one:

$$\{X, \emptyset\} \otimes \mathcal{B} \stackrel{\Delta_\pi}{\subset} \mathcal{A} \otimes \{Y, \emptyset\}. \quad (5)$$

The existence of a joining satisfying (5) is a criterion for S being a factor of T .

For more details on the results stated in this section, we refer the reader to de la Rue (2006b).

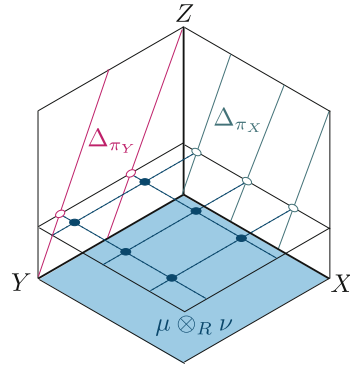
Joinings and Factors

The purpose of this section is to investigate the relationships between the disjointness of two systems S and T and the lack of a common factor. The crucial fact which was pointed out by Furstenberg is that the existence of a common factor enables one to construct a very special joining of S and T : The *relatively independent joining over this factor*.

Let us assume that our systems S and T share a common factor $(Z, \mathcal{C}, \rho, R)$, which means that we have measurable onto maps $\pi_X : X \rightarrow Z$ and $\pi_Y : Y \rightarrow Z$, respectively, sending μ and ν to ρ and satisfying $\pi_X \circ T = R \circ \pi_X$ and $\pi_Y \circ S = R \circ \pi_Y$. We can then consider the joinings supported on their graphs $\Delta_{\pi_X} \in J(T, R)$ and $\Delta_{\pi_Y} \in J(S, R)$, as defined in the preceding section. Next, we construct a joining λ of the three systems S , T , and R . Heuristically, λ is the probability distribution of the triple (x, y, z) when we first pick z according to the probability distribution ρ , then x and y according to their conditional distribution knowing z in the respective joinings Δ_{π_X} and Δ_{π_Y} , but independently of each other (Fig. 2). More precisely, λ is defined by setting, for all $A \in \mathcal{A}$, $B \in \mathcal{B}$, and $C \in \mathcal{C}$

$$\lambda(A \times B \times C) := \int_C \mathbb{E}_{\Delta_{\pi_X}} [1_{x \in A} | z] \mathbb{E}_{\Delta_{\pi_Y}} [1_{y \in B} | z] d\rho(z). \quad (6)$$

Observe that the twofold marginals of λ on $X \times Z$ and $Y \times Z$ are, respectively, Δ_{π_X} and Δ_{π_Y} , which means that we have $z = \pi_X(x) = \pi_Y(y)$ λ -almost surely.



Joinings in Ergodic Theory, Fig. 2 The relatively independent joining $\mu \otimes_R \nu$ and its disintegration over z

In other words, we have identified in the two systems T and S the projections on their common factor R . The twofold marginal of λ on $X \times Y$ is itself a joining of T and S , which we call the *relatively independent joining over the common factor R* . This joining will be denoted by $\mu \otimes_R \nu$. (Be careful: The projections π_X and π_Y are hidden in this notation, but we have to know them to define this joining.) From (6), we immediately get the formula defining $\mu \otimes_R \nu$:

$$\forall A \in \mathcal{A}, \forall B \in \mathcal{B},$$

$$\mu \otimes_R \nu(A \times B) := \int_Z \mathbb{E}_{\Delta_{\pi_X}} [1_{x \in A} | z] \mathbb{E}_{\Delta_{\pi_Y}} [1_{y \in B} | z] d\rho(z). \quad (7)$$

This definition of the relatively independent joining over a common factor can easily be extended to a finite or countable family of systems sharing the same common factor.

Note that $\mu \otimes_R \nu$ coincides with the product measure $\mu \otimes \nu$ if and only if the common factor is the trivial one-point system. We therefore get the following result:

Theorem 4

If S and T have a nontrivial common factor, then these systems are not disjoint.

As we already said in the introduction, Rudolph exhibited in Rudolph (1979) a counterexample showing that the converse is not true. There exists however an important result, which was published in Glasner et al. (2000) and Lemańczyk et al. (2000) allowing us to derive some information on factors from the non-disjointness of two systems.

Theorem 5

If T and S are not disjoint, then S has a nontrivial common factor with some joining of a countable family of copies of T .

This result leads to the introduction of a special class of factors when some dynamical system T is given: For any other dynamical system S , call T -factor of S any common factor of S with a joining of countably many copies of T . If (Z, C, ρ, R) is a T -factor of S and $\pi : Y \rightarrow Z$ is a factor map, we say that the σ -algebra $\pi^{-1}(C)$ is a T -factor σ -algebra of S . Another way to state Theorem 5 is then the following: If S and T are not disjoint, then S has a nontrivial T -factor. In fact, an even more precise result can be derived from the proof of Theorem 5: For any joining λ of S and T , for any bounded measurable function f on X , the factor σ -algebra of S generated by the function $\mathbb{E}_\lambda[f(x)|y]$ is a T -factor σ -algebra of S .

With the notion of T -factor, Theorem 5 has been extended in Lesigne et al. (2003) in the following way, showing the existence of a special T -factor σ -algebra of S comprising anything in S which could lead to a nontrivial joining between T and S .

Theorem 6

Given two measure-preserving dynamical systems $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$, there always exists a maximum T -factor σ -algebra of S , denoted by $\mathcal{F}_T$.

Under any joining λ of T and S , the σ -algebras $\mathcal{A} \otimes \{\emptyset, Y\}$ and $\{\emptyset, X\} \otimes \mathcal{B}$ are independent conditionally to the σ -algebra $\{\emptyset, X\} \otimes \mathcal{F}_T$.

Theorem 5 gives a powerful tool to prove some important disjointness results, such as those stated in Theorem 1. These results involve properties of

dynamical systems which are stable under the operations of taking joinings and factors. We will call these properties *stable properties*. This is, e.g., the case of the zero-entropy property: We know that any factor of a zero-entropy system still has zero entropy and that any joining of zero-entropy systems also has zero entropy. In other words, T has zero entropy implies that any T -factor has zero entropy. But the property of S being a K -system is precisely characterized by the fact that any nontrivial factor of S has positive entropy. Hence, a K -system S cannot have a nontrivial T -factor if T has zero entropy and is therefore disjoint from T . The converse is a consequence of Theorem 4: If S is not a K -system, then it possesses a nontrivial zero-entropy factor, and therefore there exists some zero-entropy system from which it is not disjoint.

The same argument also applies to the disjointness of discrete-spectrum systems with weakly mixing systems, since discrete spectrum is a stable property, and weakly mixing systems are characterized by the fact that they do not have any discrete-spectrum factor.

Markov Intertwinings and Composition of Joinings

There is another way of defining joinings of two measure-preserving dynamical systems involving operators on L^2 spaces, mainly put to light by Ryzhikov (see Ryzhikov 1993b): Observe that for any joining $\lambda \in \mathcal{J}(T, S)$, we can consider the operator $P_\lambda : L^2(X, \mu) \rightarrow L^2(Y, \nu)$ defined by

$$P_\lambda(f) := \mathbb{E}_\lambda[f(x)|y].$$

It is easily checked that P_λ is a Markov intertwining of T and S . Conversely, given any Markov intertwining P of T and S , it can be shown that the measure λ_P defined on $X \times Y$ by

$$\lambda_P(A \times B) := \langle P1_A, 1_B \rangle_{L^2(Y, \nu)}$$

is a joining of T and S .

This reformulation of the notion of joining is useful when joinings are studied in connection

with spectral properties of the transformations (see, e.g., Goodson 2000). It also provides us with a convenient setting to introduce the composition of joinings: If we are given three dynamical systems $(X, \mathcal{A}, \mu, T)$, $(Y, \mathcal{B}, \nu, S)$, and $(Z, \mathcal{C}, \rho, R)$, a joining $\lambda \in J(T, S)$, and a joining $\lambda' \in J(S, R)$, the composition of the Markov intertwining P_λ and $P_{\lambda'}$ is easily seen to give a third Markov intertwining, which itself corresponds to a joining of T and R denoted by $\lambda \circ \lambda'$. When $R = S = T$, i.e., when we are speaking of *twofold self-joinings* of a single system T (cf. next section), this operation turns $J(T, T) = J^2(T)$ into a semigroup. Ahn and Lemańczyk (2003) have shown that the subset $J_e^2(T)$ of ergodic twofold self-joinings is a sub-semigroup if and only if T is semisimple (see section “Simple Systems”).

Self-Joinings

We now turn to the case where the measure-preserving dynamical systems we want to join together are all copies of a single system T . For $k \geq 2$, any joining of k copies of T is called a *k-fold self-joining of T* . We denote by $J^k(T)$ the set of all k -fold self-joinings of T and by $J_e^k(T)$ the subset of *ergodic k-fold self-joinings*.

Self-Joinings and Commuting Transformations

As soon as T is not the trivial single-point system, T is never disjoint from itself: Since T is obviously isomorphic to itself, we can always find a twofold self-joining of T which is not the product measure by considering self-joinings supported on graphs of isomorphisms (see section “Joinings and Isomorphism”). The simplest of them is obtained by taking the identity map as an isomorphism, and we get that $J^2(T)$ always contains the *diagonal measure* $\Delta_0 := \Delta_{\text{Id}}$.

In general, an isomorphism of T with itself is an invertible measure-preserving transformation S of $(X, \mathcal{A}, \mu)$ which commutes with T . We call *commutant of T* the set of all such transformations (it is a subgroup of the group of automorphisms of

$(X, \mathcal{A}, \mu)$) and denote it by $C(T)$. It always contains, at least, all the powers T^n , $n \in \mathbb{Z}$.

Each element S of $C(T)$ gives rise to a twofold self-joining Δ_S supported on the graph of S . Such self-joinings are called *off-diagonal self-joinings*. They also belong to $J_e^k(T)$ if T is ergodic.

It follows that properties of the commutant of an ergodic T can be seen in its ergodic joinings. As an example of application, we can cite Ryzhikov’s proof of King’s *weak closure theorem* for rank-one transformations (An introduction to finite-rank transformations can be found, e.g., in Nadkarni (1998a); we also refer the reader to the quite complete survey (Ferenczi 1997)). Rank-one measure-preserving transformations form a very important class of zero-entropy, ergodic measure-preserving transformations. They have many remarkable properties, among which is the fact that their commutant is reduced to the weak limits of powers of T . In other words, if T is rank one, for any $S \in C(T)$, there exists a subsequence of integers (n_k) such that

$$\forall A \in \mathcal{A}, \mu(T^{-n_k} A \Delta_S^{-1} A) \xrightarrow[k \rightarrow \infty]{} 0. \quad (8)$$

King proved this result in 1986 (King 1986), using a very intricate coding argument. Observing that (8) was equivalent to the convergence, in $J^2(T)$, of $\Delta_{T^{n_k}}$ to Δ_S , Ryzhikov showed in Ryzhikov (1992b) that King’s theorem could be seen as a consequence of the following general result concerning twofold self-joinings of rank-one systems:

Theorem 7

Let T be a rank-one measure-preserving transformation, and $\lambda \in J_e^2(T)$. Then there exist $T \geq 1/2$, a subsequence of integers (n_k) , and another twofold self-joining λ' of T such that

$$\Delta_{T^{n_k}} \xrightarrow[k \rightarrow \infty]{} t\lambda + (1-t)\lambda'.$$

Minimal Self-Joinings

For any measure-preserving dynamical system T , the set of twofold self-joinings of T contains at least the product measure $\mu \otimes \mu$, the off-diagonal

joinings Δ_{T^n} for each $n \in \mathbb{Z}$, and any convex combination of these. Rudolph (1979) discovered in 1979 that we can find systems for which there are no other twofold self-joinings than these obvious ones. When this is the case, we say that T has *twofold minimal self-joinings*, or for short: $T \in \text{MSJ}(2)$. It can be shown (see, e.g., Rudolph 1990) that, as soon as the underlying probability space is not atomic (which we henceforth assume), twofold minimal self-joinings imply that T is weakly mixing and therefore that $\mu \otimes \mu$ and Δ_{T^n} , $n \in \mathbb{Z}$, are *ergodic* twofold self-joinings of T . That is why twofold minimal self-joinings are often defined by the following:

$$\begin{aligned} T \in \text{MSJ}(2) &\Leftrightarrow J_e^2(T) \\ &= \{\mu \otimes \mu\} \cup \{\Delta_{T^n}, n \in \mathbb{Z}\}. \end{aligned} \quad (9)$$

Systems with twofold minimal self-joinings have very interesting properties. First, since for any S in $C(T)$, Δ_S belongs to $J_e^2(T)$, we immediately see that the commutant of T is reduced to the powers of T . In particular, it is impossible to find a square root of T , i.e., a measure-preserving S such that $S \circ S = T$. Second, the existence of a non-trivial factor σ -algebra of T would lead, via the relatively independent self-joining over this factor, to some ergodic twofold self-joining of T which is not in the list prescribed by (9). Therefore, any factor σ -algebra of a system with twofold minimal self-joinings must be either the trivial σ -algebra $\{\emptyset, X\}$ or the whole σ -algebra $\mathcal{A}$. This has the remarkable consequence that if ξ is any random variable on the underlying probability space which is not almost-surely constant, then the process $(\xi \circ T^n)_{n \in \mathbb{Z}}$ always generates the whole σ -algebra $\mathcal{A}$. This also implies that T has zero entropy, since positive-entropy systems have many nontrivial factors.

The notion of twofold minimal self-joinings extends for any integer $k \geq 2$ to *k-fold minimal self-joinings*, which roughly means that there are no other k -fold ergodic self-joinings than the “obvious” ones: those for which the k coordinates are either independent or just translated by some power of T (see the glossary for a more precise definition). We denote in this case: $T \in \text{MSJ}(k)$. If T has k -fold

minimal self-joinings for all $k \geq 2$, we simply say that T has *minimal self-joinings*.

Rudolph’s construction of a system with twofold minimal self-joinings (Rudolph 1979) was inspired by a famous work of Ornstein (1972), giving the first example of a transformation with no roots. It turned out that Ornstein’s example is a mixing rank-one system, and all mixing rank-one systems were later proved by King (1988) to have twofold minimal self-joinings. This can also be viewed as a consequence of Ryzhikov’s Theorem 7. Indeed, in the language of joinings, the mixing property of T translates as follows:

$$T \text{ is mixing} \Leftrightarrow \Delta_{T^n} \xrightarrow[|n| \rightarrow \infty]{} \mu \otimes \mu. \quad (10)$$

Therefore, if in Theorem 7 we further assume that T is mixing, then either the sequence (n_k) we get in the conclusion is bounded, and then λ is some Δ_{T^n} , or it is unbounded and then $\lambda = \mu \otimes \mu$.

$T \in \text{MSJ}(k)$ obviously implies $T \in \text{MSJ}(k')$ for any $2 \leq k' \leq k$, but the converse is not known. The question whether twofold minimal self-joinings implies k -fold minimal self-joinings for all k is related to the important open problem of *pairwise-independent joinings* (see section “Pairwise-Independent Joinings”). But the latter problem is solved for some special classes of systems, in particular in the category of mixing rank-one transformations. It follows that, if T is mixing and rank one, then T has minimal self-joinings.

In 1980, Del Junco, Rahe, and Swanson proved that Chacon’s transformation also has minimal self-joinings (del Junco et al. 1980). This well-known transformation is also a rank-one system, but it is not mixing (it had been introduced by R.V. Chacon in 1969 (Chacon 1969) as the first explicit example of a weakly mixing transformation which is not mixing). For another example of a transformation with twofold minimal self-joinings, constructed as an exchange map on three intervals, we refer to del Junco (1983).

The existence of a transformation with minimal self-joinings has been used by Rudolph as a wonderful tool to construct a large variety of striking counterexamples, such as:

- A transformation T which has no roots, while T^2 has roots of any order
- A transformation with a cubic root but no square root
- Two measure-preserving *dynamical* systems which are weakly isomorphic (each one is a factor of the other) but not isomorphic

Let us now sketch the argument showing that we can find two systems with no common factor but which are not disjoint: We start with a system T with minimal self-joinings. Consider the direct product of T with an independent copy T' of itself, and take the *symmetric factor* S of $T \otimes T'$, that is to say the factor we get if we only look at the non-ordered pair of coordinates $\{x, x'\}$ in the Cartesian product. Then S is surely not disjoint from T , since the pair $\{x, x'\}$ is not independent of x . However, if S and T had a nontrivial common factor, then this factor should be isomorphic to T itself (because T has minimal self-joinings). Therefore we could find in the direct product $T \otimes T'$ a third copy $\tilde{T}$ of T , which is measurable with respect to the symmetric factor. In particular, $\tilde{T}$ is invariant by the flip map $(x, x') \mapsto (x', x)$, and this prevents $\tilde{T}$ from being measurable with respect to only one coordinate. Then, since $T \in \text{MSJ}(3)$, the systems T , T' , and $\tilde{T}$ have no choice but to be independent. But this contradicts the fact that $\tilde{T}$ is measurable with respect to the σ -algebra generated by x and x' . Hence, T and S have no nontrivial common factor.

We can also cite the example given by Glasner and Weiss (1983) of a pair of horocycle transformations which have no nontrivial common factor, yet are not disjoint. Their construction relies on the deep work by Ratner (1983), which describes the structure of joinings of horocycle flows.

Simple Systems

An important generalization of twofold minimal self-joinings has been proposed by William A. Veech in 1982 (Veech 1982). We say that the measure-preserving dynamical system T is *twofold simple* if it has no other ergodic twofold self-joinings than the product measure $\{\mu \otimes \mu\}$ and joinings supported on the graph of a transformation $S \in C(T)$. (The difference with $\text{MSJ}(2)$ lies in

the fact that $C(T)$ may contain other transformations than the powers of T .) It turns out that simple systems may have nontrivial factors, but the structure of these factors can be explicitly described: They are always associated with some compact subgroup of $C(T)$. More precisely, if K is a compact subgroup of $C(T)$, we can consider the factor σ -algebra

$$\mathcal{F}_K := \{A \in \mathcal{A} : \forall S \in K, A = S(A)\},$$

and the corresponding factor transformation $T|_{\mathcal{F}_K}$ (called a *group factor*). Then Veech proved the following theorem concerning the structure of factors of a twofold simple system.

Theorem 8

If the dynamical system T is twofold simple, and if $\mathcal{F} \subset \mathcal{A}$ is a nontrivial factor σ -algebra of T , then there exists a compact subgroup K of the group $C(T)$ such that $\mathcal{F} = \mathcal{F}_K$.

There is a natural generalization of Veech's property to the case of k -fold self-joinings, which has been introduced by Del Junco and Rudolph in 1987 (del Junco and Rudolph 1987) (see the precise definition of simple systems in the glossary). In their work, important results concerning the structure of factors and joinings of simple systems are proved. In particular, they are able to completely describe the structure of the ergodic joinings between a given simple system and any ergodic system (see also Glasner 2003 and Thouvenot 1995). Recall that, for any $r \geq 1$, the *symmetric factor* of $T^{\otimes r}$ is the system we get if we observe the r coordinates of the point in X^r and forget their order. This is a special case of group factor, associated with the subgroup of $C(T^{\otimes r})$ consisting of all permutations of the coordinates. We denote this symmetric factor by $T^{(r)}$.

Theorem 9

Let T be a simple system and S an ergodic system. Assume that λ is an ergodic joining of T and S which is different from the product measure. Then there exist a compact subgroup K of $C(T)$ and an integer $r \geq 1$ such that:

- $(T|_{\mathcal{F}_K})^{(r)}$ is a factor of S .
- λ is the projection on $X \times Y$ of the relatively independent joining of $T^{\otimes r}$ and S over their common factor $(T|_{\mathcal{F}_K})^{(r)}$.

If we further assume that the second system is also simple, then in the conclusion we can take $r = 1$. In other words, ergodic joinings of simple systems S and T are either the product measure or relatively independent joinings over a common group factor. This leads to the following corollary:

Theorem 10

Simple systems without nontrivial common factor are disjoint.

As for minimal self-joining, it is not known in general whether twofold simplicity implies k -fold simplicity for all k . This question is studied in Glasner et al. (1992), where sufficient spectral conditions are given for this implication to hold. It is also proved that any threefold simple weakly mixing transformation is simple of all order.

Relative Properties with Respect to a Factor

In fact, Veech also introduced a weaker, “relativized” version of the twofold simplicity. If $\mathcal{F} \subset \mathcal{A}$ is a nontrivial factor σ -algebra of T , let us denote by $J^2(T, \mathcal{F})$ the set of twofold self-joinings of T which are “constructed over $\mathcal{F}$ ”, which means that their restriction to the product σ -algebra $\mathcal{F} \otimes \mathcal{F}$ coincides with the diagonal measure. (The relatively independent joining over $\mathcal{F}$ is the canonical example of such a joining.) For the conclusion of Theorem 8 to hold, it is enough to assume only that the ergodic elements of $J^2(T, \mathcal{F})$ be supported on the graph of a transformation $S \in C(T)$. This is an important situation where the study of $J^2(T, \mathcal{F})$ gives strong informations on the way $\mathcal{F}$ is embedded in the whole system T or, in other words, on the relative properties of T with respect to the factor $T|_{\mathcal{F}}$. A simple example of such a relative property is the *relative weak mixing with respect to $\mathcal{F}$* , which is characterized by the ergodicity of the relatively independent joining

over $\mathcal{F}$ (recall that weak mixing is itself characterized by the ergodicity of the direct product $T \otimes T$).

For more details on this subject, we refer the reader to Lemańczyk et al. (2002). We also wish to mention the generalization of simplicity called *semisimplicity* proposed by Del Junco et al. in (1995), which is precisely characterized by the fact that, for any $\lambda \in J_e^2(T)$, the system $(T \otimes T)_\lambda$ is a relatively weakly mixing extension of T .

Some Applications and Future Directions

Filtering Problems

Filtering problems were one of the motivations presented by Furstenberg for the introduction of the disjointness property in Furstenberg (1967), and he considered in particular the following situation. Suppose we are given two real-valued stationary processes (X_n) and (Y_n) , with their joint distribution also stationary. We can interpret (X_n) as a signal, perturbed by a noise (Y_n) . Under which condition can we recover the original signal (X_n) from the observation of $(X_n + Y_n)$? If this is possible, we say that the sequence (X_n, Y_n) admits a *perfect filter*. Furstenberg proved that a perfect filter exists if the two processes (X_n) and (Y_n) are integrable and if the two measure-preserving dynamical systems constructed as the shift of the two processes are disjoint. Furstenberg also observed that the integrability assumption can be removed if a stronger disjointness property is satisfied: A perfect filter exists if the system T generated by (X_n) is *doubly disjoint* from the system S generated by (Y_n) , in the sense that T is disjoint from any ergodic self-joining of S . Several generalizations have been studied (see Bulatek et al. 2005 and Furstenberg et al. 1995), but the question whether the integrability assumption of the processes can be removed remained open for a long time. Finally, a short and clever argument by Garbit (2011) proved in 2011 that indeed the only assumption of disjointness is enough to ensure the existence of a perfect filter.

Joinings Proofs of Ornstein's and Krieger's Theorems

We have already seen that joinings could be used to prove isomorphisms between systems. This fact found a nice application in the proofs of two major theorems in ergodic theory: Ornstein's isomorphism theorem (Ornstein 1970), stating that two Bernoulli shifts with the same entropy are isomorphic, and Krieger's finite generator theorem (Krieger 1970), which says that any dynamical system with finite entropy is isomorphic to the shift transformation on a finite-valued stationary process. The idea of this joining approach to the proofs of Krieger's and Ornstein's theorems was originally due to Burton and Rothstein, who circulated a preliminary report on the subject which was never published (Burton and Rothstein 1977). The first published and fully detailed exposition of these proofs can be found in Rudolph's book (Rudolph 1990) (see also in Glasner's book (Glasner 2003)).

In fact, Ornstein's theorem goes far more beyond the isomorphism of two given Bernoulli shifts: It also gives a powerful tool for showing that a specific dynamical system is isomorphic to a Bernoulli shift. In particular, Ornstein introduced the property for an ergodic stationary process to be *finitely determined*. We shall not give here the precise definition of this property (for a complete exposition of Ornstein's theory, we refer the reader to Ornstein (1974)), but simply point out that Bernoulli shifts and mixing Markov chains are examples of finitely determined processes. Rudolph's argument to show Ornstein's theorem via joinings makes use of Theorem 3 and of the topology of $J(T, S)$.

Theorem 11

(Ornstein's isomorphism theorem). *Let T and S be two ergodic dynamical systems with the same entropy, and both generated by finitely determined stationary processes. Then the set of joinings of T and S which are supported on graphs of isomorphisms forms a dense G_δ in $J_e(T, S)$.*

Krieger's theorem is not as easily stated in terms of joinings, because it does not refer to the

isomorphism of two specific systems, but rather to the isomorphism of one given system with some other system which has to be found. We have therefore to introduce a larger set of joinings: Given an integer n , we denote by Y_n the set of double-sided sequences taking values in $\{1, \dots, n\}$. We consider on Y_n the shift transformation S , but we do not determine yet the invariant measure. Now, for a specific measure-preserving dynamical system T , consider the set $J(n, T)$ of all possible joinings of T with some system (Y_n, ν, S) , when ν ranges over all possible shift-invariant probability measures on Y_n . $J(n, T)$ can also be equipped with a topology which turns it into a compact convex metric space, and as soon as T is ergodic, the set $J_e(n, T)$ of ergodic elements of $J(n, T)$ is not empty. In this setting, Krieger's theorem can be stated as follows:

Theorem 12

(Krieger's finite generator theorem). *Let T be an ergodic dynamical system with entropy $h(T) < \log_2 n$. Then the set of $\lambda \in J(n, T)$ which are supported on graphs of isomorphisms between T and some system (Y_n, ν, S) forms a dense G_δ in $J_e(n, T)$.*

Since any system of the form (Y_n, ν, S) obviously has an n -valued generating process, we obtain as a corollary that T itself is generated by an n -valued process.

As another nice example of how joinings can be used to describe the isomorphisms between two transformations, we can also cite the paper (Foreman et al. 2011) by Foreman, Rudolph, and Weiss. With a clever construction of a family of transformations for which they can completely describe the joinings between T and its inverse, they are able to prove that the isomorphism relation between measure-preserving transformations of the unit interval is not Borel.

Joinings and Rohlin's Multifold-Mixing Question

We have already seen that the property of T being mixing could be expressed in terms of twofold

self-joinings of T (see (10)). Rohlin proposed in 1949 (Rohlin 1949) a generalization of this property, called *multifold mixing*: The measure-preserving transformation T is said to be *k-fold mixing* if $\forall A_1, A_2, \dots, A_k \in \mathcal{A}$,

$$\lim_{n_2, n_3, \dots, n_k \rightarrow \infty} \mu \left(A_1 \cap T^{-n_2} A_2 \cap \dots \cap T^{-(n_2 + \dots + n_k)} A_k \right) = \prod_{i=1}^k \mu(A_i).$$

Again, this definition can easily be translated into the language of joinings: T is *k-fold mixing* when the sequence $(\Delta_{T^{n_2}, \dots, T^{n_2 + \dots + n_k}})$ converges in $J^k(T)$ to $\mu^{\otimes k}$ as $n_2, \dots, n_k$ go to infinity, where $(\Delta_{T^{n_2}, \dots, T^{n_2 + \dots + n_k}})$ is the obvious generalization of Δ_{T^n} to the case of *k-fold self-joinings*. The classical notion of mixing corresponds in this setting to twofold mixing. (We must point out that Rohlin's original definition of *k-fold mixing* involved $k + 1$ sets; thus, the classical mixing property was called *onefold mixing*. However it seems that the convention we adopt here is now used by most authors, and we find it more coherent when translated in the language of multifold self-joinings.)

Obviously, threefold mixing is stronger than twofold mixing, and Rohlin asked in his article whether the converse is true. This question is still open today, even though many important works dealt with it and supplied partial answers. Most of these works directly involve self-joinings via the argument exposed in the following section.

Pairwise-Independent Joinings

Let T be a twofold mixing dynamical system. If T is not threefold mixing, $(\Delta_{T^n, T^{n+m}})$ does not converge to the product measure as n and m go to ∞ . By compactness of $J^3(T)$, we can find subsequences (n_k) and (m_k) such that $(\Delta_{T^{n_k}, T^{n_k+m_k}})$ converges to a cluster point $\lambda \neq \mu \otimes \mu$. However, by twofold mixing, the three coordinates must be pairwise independent under λ . We therefore get a threefold self-joining λ with the unusual property that λ has pairwise-independent coordinates, but λ is not the product measure.

A system is said to be *pairwise independently determined* (PID) if for all $k \geq 3$, the only pairwise-independent *k-fold self-joining* of the system is the product measure. Thus, in the class of PID systems, twofold mixing implies threefold mixing.

In fact, non-PID systems are easy to find (see, e.g., de la Rue 2006b), but the examples we know so far are either periodic transformations (which cannot be counterexamples to Rohlin's question since they are not mixing) or transformations with positive entropy. However, using an argument provided by Thouvenot, we can prove that, if there exists a twofold mixing T which is not threefold mixing, then we can find such a T in the category of zero-entropy dynamical systems (see, e.g., de la Rue 2006a). Therefore, a negative answer to the following question would solve Rohlin's multifold-mixing problem:

Question 1

Does there exist a zero-entropy, weakly mixing dynamical system T which is not PID?

The problem of (non)existence of such pairwise-independent joinings is also related to the question whether MSJ(2) implies MSJ(3) or whether twofold simplicity implies threefold simplicity. Indeed, any counterexample to one of these implications would necessarily be of zero entropy and would possess a pairwise-independent threefold self-joining which is not the product measure. Moreover, it is noticed in [24, Corollary 12.22] that if we find a system which is MSJ(2) but not MSJ(3), then this system would also be twofold mixing without being threefold mixing.

Question 1 has been answered by Host and Ryzhikov for some special classes of zero-entropy dynamical systems.

Host's and Ryzhikov's Theorems

The following theorem, proved in 1991 by Host (1991) (see also Glasner 2003 and Nadkarni 1998b), establishes a spectacular connection between the spectral properties of a finite family of dynamical systems and the nonexistence

of a pairwise-independent, nonindependent joining:

Theorem 13

(Host's theorem on singular spectrum). *Let $(X_i, \mathcal{A}_i, \mu_i, T_i)_{1 \leq i \leq r}$ be a finite family of measure-preserving dynamical systems with purely singular spectrum. Then any pairwise-independent joining $\lambda \in J(T_1, \dots, T_r)$ is the product measure $\mu_1 \otimes \dots \otimes \mu_r$.*

Corollary 1

If a dynamical system with singular spectrum is twofold mixing, then it is k -fold mixing for any $k \geq 2$.

The multifold-mixing problem for rank-one measure-preserving systems was solved in 1984 by Kalikow (1984), using arguments which do not involve the theory of joinings. In 1993, Ryzhikov (Ryzhikov 1993a) extended Kalikow's result to finite-rank systems, by giving a negative answer to Question 1 in the category of finite-rank mixing systems:

Theorem 14

(Ryzhikov's theorem for finite-rank systems). *All finite-rank mixing transformations are PID.*

Corollary 2

If a finite-rank transformation is twofold mixing, then it is k -fold mixing for any $k \geq 2$.

Joinings and Multiple Ergodic Averages

Important researches in ergodic theory are devoted to the convergence of so-called multiple ergodic averages, that is to say expressions of the form

$$\frac{1}{n} \sum_{k=0}^{n-1} f_1(T_1^k x) f_2(T_2^k x) \cdots f_d(T_d^k x), \quad (11)$$

where $d \geq 2$, $T_1, \dots, T_d$ are d measure-preserving transformations of the probability space $(X, \mathcal{A}, \mu)$

and $f_1, \dots, f_d$ are bounded measurable functions. A case of special interest is when $T_i = T^i (i = 1, \dots, d)$ for some measure-preserving transformation T , which is strongly connected to Furstenberg's ergodic proof of Szemerédi's theorem, and more generally when $T_1, \dots, T_d$ are commuting transformations.

The study of convergence in L^2 of (11) has a long history and has led to the emergence of deep and successful ideas. We can cite in particular the work of Host and Kra (2005) who established the L^2 -convergence for successive powers of a single transformation, introducing and describing the structure of a sequence of factors for which joinings play a central role. The first proof of the convergence in L^2 of (11) in the case of d commuting transformations has been given by Tao in (2008). While Tao's original proof relies on a purely combinatorial argument and does not make use of joinings, we would like here to emphasize the alternative proof established by Austin (2010), which strongly relies on the theory of joinings, and how they are involved in the description of the structure of the factors of a dynamical system. For an even more joining oriented version of Austin's proof, see de la Rue (2009).

As far as pointwise convergence of multiple ergodic averages is concerned, still more questions remain open, and they certainly also involve the theory of joinings. As a simple example, we present the relationships between the study of pointwise convergence of (11) when $d = 2$, that is, we consider two nonnecessarily commuting transformations S and T and we study the sequence

$$\left(\frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) g(S^k x) \right)_{n>0}. \quad (12)$$

It turns out that disjointness of T and S is a sufficient condition for the almost-sure convergence to hold. Indeed, let us first consider the case where T and S are defined on a priori different spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$, respectively, and consider the ergodic average in the product

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) g(S^k y), \quad (13)$$

which can be viewed as the integral of the function $f \otimes g$ with respect to the empirical distribution

$$\delta_n(x, y) := \frac{1}{n} \sum_{k=0}^{n-1} \delta_{(T^k x, S^k y)}.$$

We can always assume that T and S are continuous transformations of compact metric spaces (indeed, any measure-preserving dynamical system is isomorphic to such a transformation on a compact metric space: see, e.g., Furstenberg 1981). Then the set of probability measures on $X \times Y$ equipped with the topology of weak convergence is metric compact. Now, here is the crucial point where joinings appear: If T and S are ergodic, we can easily find subsets $X_0 \subset X$ and $Y_0 \subset Y$ with $\mu(X_0) = \nu(Y_0) = 1$, such that for all $(x, y) \in X_0 \times Y_0$, any cluster point of the sequence $(\delta_n(x, y))_{n>0}$ is automatically a joining of T and S . (We just have to pick x and y among the “good points” for the ergodic theorem in their respective spaces.) When T and S are disjoint, the only cluster point to the sequence $(\delta_n(x, y))$ is therefore $\mu \otimes \nu$. This ensures that, for continuous f and g , (13) converges to the product of the integrals of f and g as soon as (x, y) is picked in $X_0 \times Y_0$. The subspace of continuous functions being dense in L^2 , the classical ergodic maximal inequality (see Garsia 1970) ensures that, for any f and g in $L^2(\mu)$, (13) converges for any (x, y) in a rectangle of full measure $X_0 \times Y_0$.

Coming back to the original question where the spaces on which T and S act are identified, we observe that with probability one, x belongs both to X_0 and Y_0 , and therefore the sequence (12) converges.

The existence of a rectangle of full measure in which the sequence of empirical distributions $(\delta_n(x, y))_{n>0}$ always converges to some joining has been studied in Lesigne et al. (2003) as a natural generalization of the notion of disjointness. This property was called *weak disjointness* of S and T , and it is indeed strictly weaker than disjointness,

since there are examples of transformations which are weakly disjoint from themselves.

Recent works use joinings to study the pointwise convergence of multiple ergodic averages. We mention in particular Huang et al. (2014), where the case of successive powers of an ergodic distal transformation is treated, and Donoso and Sun (2016) where this result is extended to the case of d commuting transformations generating a distal action of $\mathbb{Z}^d$, using some of Austin’s ideas.

Also, let us observe that a strong connection with Question 1 on pairwise-independent joinings has been established in Gutman et al. (2018). Indeed, using a result of Bourgain according to which the almost-sure convergence of (11) holds when $d = 2$ and T_1, T_2 are powers of a single transformation, the authors prove that if T is PID and weakly mixing, then (11) converge almost surely for $T_i = T^i, i = 1, \dots, d$.

Joinings and Conjectures in Number Theory

Joinings have recently proved to be a relevant tool in the study of a conjecture by Sarnak, which lies at the interface between number theory and dynamical systems. Sarnak conjecture deals with the famous arithmetic Möbius function μ , which is defined for each integer $n \geq 0$ by

$$\mu(n) := \begin{cases} 0 & \text{if there exists a prime } p \text{ such that } p^2 \mid n, \\ (-1)^k & \text{if } n \text{ is the product of } k \text{ distinct prime numbers } (k \geq 0). \end{cases}$$

A famous heuristic called the *Möbius randomness principle* states that the patterns of symbols $-1, 1$, and 0 in the sequence μ behave so chaotically that μ has no correlation with any *reasonably simple* sequence ξ . Sarnak proposed a precise interpretation of this principle in the context of dynamical systems:

Conjecture 1

(Sarnak’s conjecture). *For any topological dynamical system (X, T) with zero topological entropy, any continuous real-valued function f defined on X , and any $x \in X$, we have*

$$\frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) \mu(k) \xrightarrow{n \rightarrow \infty} 0. \quad (14)$$

Sarnak pointed out that this conjecture was supported by an older conjecture of Chowla which says that μ has no autocorrelation of any order. More precisely, it can be stated as follows:

Conjecture 2

(Chowla's conjecture). *For any integers $r \geq 0, i_0, i_1, \dots, i_r \geq 0$ with at least one odd i_j , we have*

$$\frac{1}{n} \sum_{k=0}^{n-1} \mu(k)^{i_0} \mu(k+1)^{i_1} \cdots \mu(k+r)^{i_r} \xrightarrow{n \rightarrow \infty} 0.$$

As pointed out by Sarnak, that the Chowla conjecture implies the Sarnak conjecture can be proved via ergodic theory arguments. Indeed, an interpretation of the Chowla conjecture is that μ is the product of μ^2 (the characteristic function of the set of square-free numbers) with a sequence $\pi \in \{-1, 1\}^{\mathbb{N}}$ behaving as a typical output of an infinite sequence of balanced coin tosses. Now, with the assumptions of the statement of the Sarnak conjecture, the sequence $f(T^k x) \mu^2(k)$ is produced by a zero-entropy system, whereas the sequence π comes from a K -system. Hence (14) can be viewed as a consequence of the disjointness between K -systems and systems of zero entropy. (See, e.g., El Abdalaoui et al. (2017) for details.)

Another reason why joinings are used in the study of Sarnak conjecture comes from the following criterion introduced by Bourgain et al. (2013), which provides a sufficient condition for a bounded sequence (a_n) to be orthogonal to any bounded *multiplicative* function v . Recall that the Möbius function is itself multiplicative, which means that $\mu(mn) = \mu(m)\mu(n)$ whenever m, n are coprime integers.

Lemma 1

(Criterion of Bourgain, Sarnak, and Ziegler). *Assume that (a_n) is a bounded sequence of complex numbers, such that*

$$\limsup_{r, s \rightarrow \infty, \substack{r, s \text{ different primes}}} \limsup_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{k=0}^{n-1} a_{kr} \bar{a}_{ks} \right| = 0. \quad (15)$$

Then, for each bounded multiplicative function $v : \mathbb{N} \rightarrow \mathbb{C}$, we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} a_k \cdot v(k) = 0.$$

To apply the above lemma in the context of the Sarnak conjecture, we have to consider the sequence $a_k = f(T^k x)$; hence (15) becomes an assumption on this kind of multiple ergodic average

$$\frac{1}{n} \sum_{k=0}^{n-1} f((T^r)^k x) \overline{f((T^s)^k x)}.$$

Now, let us assume for simplification that the topological dynamical system (X, T) has a unique invariant probability measure μ . Then arguments similar to those exposed in section “[Joinings and Multiple Ergodic Averages](#)” lead to considering joinings of different prime powers T^r and T^s of the measure-preserving system (X, T, μ) . For example, as explained in Bourgain et al. (2013), disjointness of different prime powers of T implies the validity of the Sarnak conjecture for the system. But Lemma 1 can also be applied in a larger class of systems, which is defined below and for which we can control all possible joinings of different prime powers of T .

The ergodic measure-preserving system (X, μ, T) is said to have *Asymptotically Orthogonal Powers* (AOP) if for each given f and g in $L^2(\mu)$ with $\int_X f d\mu = \int_X g d\mu = 0$, we have

$$\lim_{r, s \rightarrow \infty, \substack{r, s \text{ different primes}}} \sup_{\kappa \in J_e(T^r, T^s)} \left| \int_{X \times X} f \otimes g d\kappa \right| = 0.$$

Surprisingly, the class of AOP systems includes examples where all positive powers of T are isomorphic. Examples of system with AOP and consequences of this property are given in

Ferenczi et al. (2017) and references therein. In particular, the Sarnak conjecture holds for any uniquely ergodic model of an AOP system, with uniform convergence of (14) with respect to x .

Future Directions

A lot of important open questions in ergodic theory involve joinings, and we already have cited several of them: joinings are a natural tool when we want to deal with some problems of pointwise convergence involving several transformations (see section “Joinings and Multiple Ergodic Averages”). It can therefore be assumed that they will play an important role in future progress on pointwise convergence of multiple ergodic averages. Their use is also fundamental in the study of Rohlin’s question on multifold mixing. As far as this latter problem is concerned, we may mention a recent approach to Question 1: start with a transformation for which some special pairwise-independent self-joining exists, and see what this assumption entails. In particular, we can ask under which conditions there exists a pairwise-independent threefold self-joining of T under which the third coordinate is a function of the two others. It has already been proved in Janvresse and de la Rue (2007) that if this function is sufficiently regular (continuous for some topology), then T is periodic or has positive entropy. And there are strong evidences leading to the conjecture that, when T is weakly mixing, such a situation can only arise when T is a Bernoulli shift of entropy $\log n$ for some integer $n \geq 2$. A question in the same spirit was raised by Ryzhikov, who asked in Ryzhikov (1992a) under which conditions we can find a factor of the direct product $T \times T$ which is independent of both coordinates.

There is also a lot of work to do with joinings in order to understand the structure of factors of some dynamical systems and how different classes of systems are related. An example of such a work is given in the class of Gaussian dynamical systems, i.e., dynamical systems constructed from the shift on a stationary Gaussian process: For some of them (which are called *GAG*, from the French *Gaussien à Autocouplages Gaussiens*), it can be proved that any ergodic self-joining is itself

a Gaussian system (see Lemańczyk et al. 2000 and Thouvenot 1987), and this gives a complete description of the structure of their factors. This kind of analysis is expected to be applicable to other classes of dynamical systems. In particular, Gaussian joinings find a nice generalization in the notion of *infinitely divisible joinings*, studied by Roy in (2009). These ID joinings concern a wider class of dynamical systems of probabilistic origin, among which we can also find Poisson suspensions. The counterpart of Gaussian joinings in this latter class is *Poisson joinings*, which have been introduced by Derriennic et al. As far as Poisson suspensions are concerned, the analog of the GAG property in the Gaussian class can also be considered, and examples of Poisson suspensions for which the only ergodic self-joinings are Poisson joinings have been given in Janvresse et al. (2017) and Parreau and Roy (2007). In Derriennic et al. a general joining property is described: T satisfies the *ELF property* (from the French *Ergodicité des Limites Faibles*) if any joining which can be obtained as a limit of off-diagonal joinings $\Delta_{T^{n_k}}$ is automatically ergodic. It turns out that this property is satisfied by any system arising from an infinitely divisible stationary process (see Derriennic et al. and Roy 2009). It is proved in Derriennic et al. that the ELF property implies disjointness with any system which is twofold simple and weakly mixing but not mixing. The ELF property is expected to give a useful tool to prove disjointness between dynamical systems of probabilistic origin and other classes of systems (see, e.g., Frączek and Lemańczyk 2004) in the case of $\mathbb{R}$ -action for disjointness between ELF systems and a class of special flows over irrational rotations).

Bibliography

- Ahn Y-H, Lemańczyk M (2003) An algebraic property of joinings. Proc Am Math Soc 131(6):1711–1716. (Electronic)
- Austin T (2010) Multiple recurrence and the structure of probability-preserving systems. arXiv:1006.0491
- Bourgain J, Sarnak P, Ziegler T (2013) Disjointness of Mobius from horocycle flows. In: From Fourier analysis and number theory to Radon transforms and

- geometry, *Developments in mathematics*, vol 28. Springer, New York, pp 67–83
- Bulatek W, Lemańczyk M, Lesigne E (2005) *IEEE Trans Inform Theory* 51(10):3586–3593
- Burton R, Rothstein A (1977) Isomorphism theorems in ergodic theory. Technical report, Oregon State University
- Chacon RV (1969) Weakly mixing transformations which are not strongly mixing. *Proc Am Math Soc* 22:559–562
- de la Rue T (2006a) 2-fold and 3-fold mixing: why 3-dot-type counterexamples are impossible in one dimension. *Bull Braz Math Soc (N.S.)* 37(4):503–521
- de la Rue T (2006b) An introduction to joinings in ergodic theory. *Discret Contin Dyn Syst* 15(1):121–142
- de la Rue T (2009) Notes on Austin’s multiple ergodic theorem, hal-00400975
- del Junco A (1983) A family of counterexamples in ergodic theory. *Israel J Math* 44(2):160–188
- del Junco A, Rudolph DJ (1987) On ergodic actions whose self-joinings are graphs. *Ergodic Theory Dyn Syst* 7(4):531–557
- del Junco A, Rahe M, Swanson L (1980) Chacon’s automorphism has minimal self-joinings. *J Anal Math* 37:276–284
- del Junco A, Lemańczyk M, Mentzen MK (1995) Semi-simplicity, joinings and group extensions. *Stud Math* 112(2):141–164
- Derriennic Y, Frączek K, Lemańczyk M, Parreau F. (2008) Ergodic automorphisms whose weak closure of off-diagonal measures consists of ergodic self-joinings. To appear in *Colloquium Mathematicum* 110, No. 1, 81–115
- Donoso S, Sun W (2016) Pointwise convergence of some multiple ergodic averages. arXiv:1609.02529
- El Abdalaoui H, Kulaga-Przymus J, Lemańczyk M, de la Rue T (2017) The Chowla and the Sarnak conjectures from ergodic theory point of view. *Discret Continuous Dyn Syst* 37(6):2899–2944
- Ferenczi S (1997) Systems of finite rank. *Colloq Math* 73(1):35–65
- Ferenczi S, Kulaga-Przymus J, Lemańczyk M (2017) Sarnak’s conjecture – what’s new. arXiv:1710.04039
- Foreman M, Rudolph DJ, Weiss B (2011) The conjugacy problem in ergodic theory. *Ann Math* 173(3):1529–1586. (English)
- Frączek K, Lemańczyk M (2004) A class of special flows over irrational rotations which is disjoint from mixing flows. *Ergodic Theory Dyn Syst* 24(4):1083–1095
- Furstenberg H (1967) Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math Syst Theory* 1:1–49
- Furstenberg H (1981) Recurrence in ergodic theory and combinatorial number theory, M. B. Porter Lectures. Princeton University Press, Princeton
- Furstenberg H, Keynes HB, Shapiro L (1973) Prime flows in topological dynamics. *Isr J Math* 14:26–38. (English)
- Furstenberg H, Peres Y, Weiss B (1995) Perfect filtering and double disjointness. *Ann Inst Henri Poincaré Probab Stat* 31(3):453–465
- Garbit R (2011) A note on Furstenberg’s filtering problem. *Isr J Math* 182:333–336
- Garsia AM (1970) Topics in almost everywhere convergence, *Lectures in advanced mathematics*, vol 4. Markham Publishing Co., Chicago
- Glasner E (2003) Ergodic theory via joinings, *Mathematical surveys and monographs*, vol 101. American Mathematical Society, Providence
- Glasner S, Weiss B (1983) Minimal transformations with no common factor need not be disjoint. *Isr J Math* 45(1):1–8
- Glasner E, Host B, Rudolph DJ (1992) Simple systems and their higher order self-joinings. *Isr J Math* 78(1):131–142
- Glasner E, Thouvenot J-P, Weiss B (2000) Entropy theory without a past. *Ergodic Theory Dyn Syst* 20(5):1355–1370
- Goodson GR (2000) Joining properties of ergodic dynamical systems having simple spectrum. *Sankhyā Ser A* 62(3):307–317. Ergodic theory and harmonic analysis (Mumbai, 1999)
- Gutman Y, Huang W, Shao S, Ye X (2018) Almost sure convergence of the multiple ergodic average for certain weakly mixing systems. *Acta Math Sin Engl Ser* 34(1):79–90
- Host B (1991) Mixing of all orders and pairwise independent joinings of systems with singular spectrum. *Isr J Math* 76(3):289–298
- Host B, Kra B (2005) Nonconventional ergodic averages and nil-manifolds. *Ann Math* 161(1):397–488
- Huang W, Shao S, Ye X (2014) Pointwise convergence of multiple ergodic averages and strictly ergodic models. arXiv:1406.5930
- Janvresse É, de la Rue T (2007) On a class of pairwise-independent joinings, Preprint. arXiv:0704.3358v2 [math.PR]
- Janvresse É, Roy E, de la Rue T (2017) Poisson suspensions and Sushis. *Ann Sci Éc Norm Super* 50(6):1301–1334
- Kalikow SA (1984) Twofold mixing implies threefold mixing for rank one transformations. *Ergodic Theory Dyn Syst* 4(2):237–259
- King J (1986) The commutant is the weak closure of the powers, for rank-1 transformations. *Ergodic Theory Dyn Syst* 6(3):363–384
- King J (1988) Joining-rank and the structure of finite rank mixing transformations. *J Anal Math* 51:182–227
- King JL (1990) A map with topological minimal self-joinings in the sense of del Junco. *Ergodic Theory Dyn Syst* 10(4):745–761. (English)
- Krieger W (1970) On entropy and generators of measure-preserving transformations. *Trans Am Math Soc* 149:453–464
- Lemańczyk M, Parreau F, Thouvenot J-P (2000) Gaussian automorphisms whose ergodic self-joinings are Gaussian. *Fundam Math* 164(3):253–293

- Lemańczyk M, Thouvenot J-P, Weiss B (2002) Relative discrete spectrum and joinings. *Monatsh Math* 137(1):57–75
- Lesigne E, Rittaud B, de la Rue T (2003) Weak disjointness of measure preserving dynamical systems. *Ergodic Theory Dyn Syst* 23(4):1173–1198
- Nadkarni MG (1998a) Basic ergodic theory, Birkhäuser advanced texts: Basler Lehrbücher. [Birkhäuser advanced texts: Basel textbooks], 2nd edn. Birkhäuser Verlag, Basel
- Nadkarni MG (1998b) Spectral theory of dynamical systems, Birkhäuser advanced texts: Basler Lehrbücher. [Birkhäuser advanced texts: Basel textbooks]. Birkhäuser Verlag, Basel
- Ornstein DS (1970) Bernoulli shifts with the same entropy are isomorphic. *Adv Math* 4:337–352. 1970
- Ornstein DS (1972) On the root problem in ergodic theory. In: Proceedings of the sixth Berkeley symposium on mathematical statistics and probability (University of California, Berkeley, 1970/1971), vol II, probability theory. University of California Press, Berkeley, pp 347–356
- Ornstein DS (1974) Ergodic theory, randomness, and dynamical systems, James K. Whittemore lectures in mathematics given at Yale University, Yale mathematical monographs, vol 5. Yale University Press, New Haven
- Parreau F, Roy E (2007) Poisson suspensions with a minimal set of self-joinings, Preprint
- Ratner M (1983) Horocycle flows, joinings and rigidity of products. *Ann Math* 118(2):277–313
- Rohlin VA (1949) On endomorphisms of compact commutative groups. *Izvestiya Akad Nauk SSSR Ser Mat* 13:329–340
- Roy E (2009) Poisson suspensions and infinite ergodic theory. *Ergodic Theory Dyn Syst* 29(2):667–683
- Rudolph DJ (1979) An example of a measure preserving map with minimal self-joinings, and applications. *J Anal Math* 35:97–122
- Rudolph DJ (1990) Fundamentals of measurable dynamics: ergodic theory on Lebesgue spaces, Oxford Science Publications. The Clarendon Press/Oxford University Press, New York
- Ryzhikov VV (1992a) Stochastic wreath products and joinings of dynamical systems. *Mat Zametki* 52(3):130–140. 160
- Ryzhikov VV (1992b) Mixing, rank and minimal self-joining of actions with invariant measure. *Mat Sb* 183(3):133–160
- Ryzhikov VV (1993a) Joinings and multiple mixing of the actions of finite rank. *Funktsional Anal i Prilozhen* 27(2):63–78. 96
- Ryzhikov VV (1993b) Joinings, wreath products, factors and mixing properties of dynamical systems. *Izv Ross Akad Nauk Ser Mat* 57(1):102–128
- Sarnak P (2012) Mobius randomness and dynamics. *Not S Afr Math Soc* 43(2):89–97
- Tao T (2008) Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory Dyn Syst* 28(2):657–688
- Thorisson H (2000) Coupling, stationarity, and regeneration, Probability and its applications. Springer, New York
- Thouvenot J-P (1987) The metrical structure of some Gaussian processes. In: Proceedings of the conference on ergodic theory and related topics, II (Georgenthal, 1986), Teubner-Texte zur mathematik, vol 94. Teubner, Leipzig, pp 195–198
- Thouvenot J-P (1995) Some properties and applications of joinings in ergodic theory. In: Ergodic theory and its connections with harmonic analysis (Alexandria, 1993), London mathematical society lecture note series, vol 205. Cambridge University Press, Cambridge, pp 207–235
- Veech WA (1982) A criterion for a process to be prime. *Monatsh Math* 94(4):335–341
- Weiss B (1998) Multiple recurrence and doubly minimal systems, topological dynamics and applications. A volume in honor of Robert Ellis. In: Proceedings of a conference in honor of the retirement of Robert Ellis, Minneapolis, 5–6 Apr 1995. American Mathematical Society, Providence, pp 189–196. (English)

Entropy in Ergodic Theory

Jonathan L. F. King¹ and Kyewon Koh Park²

¹University of Florida, Gainesville, FL, USA

²Korea Institute for Advanced Study, Seoul, South Korea

Article Outline

Glossary

Definition of the Subject

Entropy Example: *How Many Questions?*

Distribution Entropy

A Glance at Shannon's Noisy Channel Theorem

The Information Function

Entropy of a Transformation

Determinism and Zero-Entropy

The Pinsker-Field and K-Automorphisms

Ornstein Theory

Skew Product and Random Walk in Random Scenery

Topological Entropy

Entropy of a Flow($\mathbb{R}$ -Action)

Entropy of $\mathbb{Z}^d$ -Actions

Finitely Observable Invariant

Recent Progress

Exodos

Bibliography

Glossary

This entry includes some easy proofs and auxiliary results which illustrate how the ideas evolved in the field. This can be also useful for persons not in the field of ergodic theory. However, due to the page constraint, we jump to the brief presentation of recent developments of entropy theory, skipping many important intermediate advances.

Some of the following definitions refer to the "Notation" paragraph immediately below. Use *mpt* for "measure-preserving transformation."

Almost everywhere (a.e.) A measure-theoretic statement holds *almost everywhere*, abbreviated *a.e.*, if it holds off of a nullset. (Eugene Gutkin once remarked to us that the problem with Measure Theory is... that you have to say "almost everywhere," almost everywhere.) For example, $B \stackrel{a.e.}{\supset} A$ means that $\mu(A \setminus B)$ is zero. The *a.e.* will usually be implicit.

Factor map A factor map $\psi : (X, \mathcal{X}, \mu, T) \rightarrow (Y, \mathcal{Y}, \nu, S)$ is a measure-preserving map $\psi : X \rightarrow Y$ which intertwines the transformations, $\psi \circ T = S \circ \psi$ after deleting a nullset (a mass-zero set) in each space. And ψ is an *isomorphism* if this ψ is a bijection and ψ^{-1} is also a factor map. If X and Y are topological spaces, then we require ψ to be continuous and onto.

Fonts We use the font $\mathcal{H}$, h , $\mathcal{I}$ for *distribution-entropy*, *entropy*, and the *information function*. In contrast, the script font $\mathcal{A}, \mathcal{B}, \mathcal{C} \dots$ will be used for collections of sets; usually subfields of $\mathcal{X}$. Use $\mathbb{E}(\cdot)$ for the (conditional) expectation operator.

Measure-preserving map A *measure-preserving map* $\psi : (X, \mathcal{X}, \mu) \rightarrow (Y, \mathcal{Y}, \nu)$ is a map $\psi : X \rightarrow Y$ such that the inverse image of each $B \in \mathcal{Y}$ is in $\mathcal{X}$, and $\mu(\psi^{-1}(B)) = \nu(B)$. A (*measure-preserving*) *transformation* is a measure-preserving map $T : (X, \mathcal{X}, \mu) \rightarrow (X, \mathcal{X}, \mu)$. Condense this notation to $(X, \mathcal{X}, \mu, T)$ or (X, μ, T) .

Measure space A *measure space* $(X, \mathcal{X}, \mu)$ is a set X , a *field* (i.e., a σ -algebra) $\mathcal{X}$ of subsets of X , and a countably-additive measure $\mu : \mathcal{X} \rightarrow [0, \infty]$. (We often just write (X, μ) , with the field implicit.) For a collection $C \subset \mathcal{X}$, use $\text{Fld}(C) \supset C$ for the smallest field including C . The number " $\mu(B)$ is the *μ -mass of B* ."

Notation $\mathbb{Z}$ = integers, $\mathbb{Z}_+$ = positive integers, and $\mathbb{N}$ = natural numbers =0, 1, 2, ... (Some well-meaning folk use $\mathbb{N}$ for $\mathbb{Z}_+$, saying "Nothing could be more natural than the positive integers." And this is why $0 \in \mathbb{N}$.) Use $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ for the *ceiling* and *floor* functions; $\lfloor \cdot \rfloor$ is also called the "*greatest-integer function*." For an interval

$J := [a, b) \subset [-\infty, +\infty]$, let $[a \dots b)$ denote the **interval of integers** $J \cap \mathbb{Z}$ (with a similar convention for closed and open intervals). For example, $(e \dots \pi) = (e \dots \pi) = \{3\}$. For subsets A and B of the same space, Ω , use $A \subset B$ for inclusion and $A \subsetneq B$ for **proper inclusion**. The difference set $B \setminus A$ is $\{\omega \in B \mid \omega \notin A\}$. Employ A^c for the complement $\Omega \setminus A$. Since we work in a probability space, if we let $x := \mu(A)$, then a convenient convention is to have x^c denote $1 - x$, since then $\mu(A^c)$ equals x^c . Use $A \Delta B$ for the **symmetric difference** $[A \setminus B] \cup [B \setminus A]$. For a collection $C = \{E_j\}_j$ of sets in Ω , let the **disjoint union** $\bigsqcup_j E_j$ or $\bigsqcup(C)$ represent the union $\cup_j E_j$ and also assert that the sets are pairwise disjoint. Use “ $\forall_{\text{large } n}$ ” to mean: “ $\exists n_0$ such that $\forall n > n_0$.” To refer to left hand side of an (20), use LhS(20); do analogously for RhS(20), the right hand side.

Partition A **partition** $P = (A_1, A_2, \dots)$ splits X into pairwise disjoint subsets $A_i \in X$ so that the disjoint union $\bigsqcup_i A_i$ is all of X . Each A_i is an **atom** of P . Use $|P|$ or $\#P$ for the number of atoms. When P partitions a probability space, then it yields a probability vector $\vec{v}$, where $v_j := \mu(A_j)$. Lastly, use $P\langle x \rangle$ to denote the P -atom that owns x .

Probability space A **probability space** is a measure space (X, μ) with $\mu(X) = 1$; this μ is a **probability measure**. All our maps/transformations in this entry are on probability spaces.

Probability vector A **probability vector** $\vec{v} = (v_1, v_2, \dots)$ is a list of nonnegative reals whose sum is 1. We generally assume that probability vectors and partitions (see below) have **finitely** many components. Write “*countable probability vector/partition*,” when finitely or denumerably many components are considered.

Topological map A topological space $(X, \mathcal{V}, T)$ is a compact set X with an open cover $\mathcal{V} = \{U_i\}$. Each U_i is open and $\cup U_i = X$. A transformation T is a continuous map from X to itself.

Definition of the Subject

The word “*entropy*” (originally German, *Entropie*) was coined by Rudolf Julius Emanuel

Clausius circa 1865 (Clausius 1864, 1867), taken from the Greek $\epsilon\nu\tau\rho o\pi\iota\alpha$, “*a turning towards*.” This entry thus begins (*Prolegomenon*, “introduction”) and ends (*Exdos* (This is the Greek spelling.), “the path out”) in Greek.

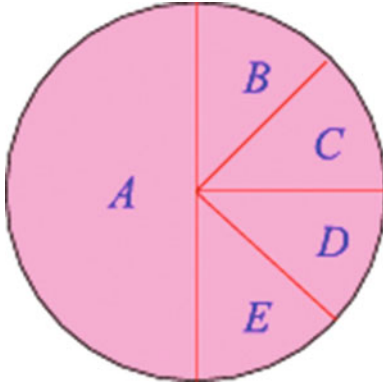
Clausius, in his coinage, was referring to the thermodynamic notion in physics. Our focus in this entry, however, will be the concept in measurable and topological dynamics. (Entropy in differentiable dynamics would require an article by itself. For instance, see Katok and Hasselblatt 1995; Ledrappier and Young 1985; Mane 1987; Pesin 1977.) Shannon’s 1948 paper (Shannon 1948) on Information Theory, then Kolmogorov’s (1958) and Sinai’s (1959) generalization to dynamical systems, will be our starting point. Our discussion will be of mostly the one-dimensional case, where the acting-group is $\mathbb{Z}$. We will briefly mention continuous time flow, $\mathbb{R}$ -action. Later we will discuss actions of group $\mathbb{R}^d$ which is a stepping stone to amenable group actions.

My greatest concern was what to call it. I thought of calling it ‘*information*’, but the word was overly used, so I decided to call it ‘*uncertainty*’. John von Neumann had a better idea, he told me, ‘You should call it entropy, for two reasons. In the first place your uncertainty function goes by that name in statistical mechanics. In the second place, and more important, *nobody knows what entropy really is*, so in a debate you will always have the advantage.’ (Shannon as quoted in Tribus and McIrvine 1971)

Entropy Example: How Many Questions?

Imagine a dartboard, Fig. 1¹, split in five regions $A, \dots, E$ with known probabilities. Blindfolded, you throw a dart at the board. What is the expected

¹In this entry, unmarked logs will be to base-2. In entropy theory, it does not matter much what base is used, but base-2 is convenient for computing entropy for messages described in bits. When using the natural logarithm, some people refer to the unit of information as a nat. In this entry, we have picked bits rather than nats. This holds when each probability p is a reciprocal power of two. For general probabilities, the “expected number of questions” interpretation holds in a weaker sense: Throw N darts independently at N copies of the dartboard. Efficiently ask Yes/No questions to determine where all N darts landed. Dividing by N , then sending $N \rightarrow \infty$, will be the $p \log\left(\frac{1}{p}\right)$ sum of (1).



Entropy in Ergodic Theory, Fig. 1 This dartboard is a probability space with a 5-set partition. The atoms have probabilities $\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}$. This probability distribution will be used later in Meshalkin's example

number V of Yes/No questions needed to ascertain the region in which the dart landed?

Solve this by always dividing the remaining probability in half. “Is it A?” if Yes, then $V = 1$. Else: “Is it B or C?” – if Yes, then “Is it B?” – if No, then the dart landed in C, and $V = 3$ was the number of questions. Evidently $V = 3$ also for regions B, D, E. Using “log” to denote base-2 logarithm, the expected number of questions is thus

$$\begin{aligned} \mathbb{E}(V) &= \frac{1}{2} \cdot 1 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 \\ &= \sum_{j=0}^4 p_j \log \left(\frac{1}{p_j} \right) \stackrel{\text{note}}{=} 2. \end{aligned} \quad (1)$$

Letting $\vec{v} := (\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$ be the probability vector, we can write this expectation as

$$\mathbb{E}(V) = \sum_{x \in \vec{v}} \eta(x).$$

Here,

$\eta: [0, 1] \rightarrow [0, \infty)$ is the important function (There does not seem to be a standard name for this function. We use η , since an uppercase η looks like an H, which is the letter that Shannon used to denote what we are calling distribution-entropy.)

$\eta(x) := x \log(1/x)$, so extending by continuity gives $\eta(0) = 0$.

(2)

An interpretation of “ $\eta(x)$ ” is the number of questions needed to winnow down to an event of probability x .

Distribution Entropy

Given a probability vector $\vec{v}$, define its *distribution entropy* as

$$\mathcal{H}(\vec{v}) := \sum_{x \in \vec{v}} \eta(x). \quad (3)$$

This entry will use the term *distropy* for “*distribution entropy*,” reserving the word *entropy* for the corresponding dynamical concept, when there is a notion of *time* involved. Getting ahead of ourselves, the *entropy* of a stationary process is the asymptotic average value that its distropy decays to, as we look at larger and larger finite portions of the process.

An equi-probable vector $\vec{v} := (\frac{1}{K}, \dots, \frac{1}{K})$ evidently has $\mathcal{H}(\vec{v}) = \log(K)$. On a probability space, the “*distropy of partition P*,” written $\mathcal{H}(P)$ or $\mathcal{H}(A_1, A_2, \dots)$ shall mean the distropy of probability vector $\vec{v} = (\mu(A_1), \mu(A_2), \dots)$.

A (finite) partition necessarily has finite distropy. A *countable* partition can have finite distropy, for example, $\mathcal{H}(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots) = 2$. One could also have infinite distropy: Consider a piece $B \subset X$ of mass $1/2^N$. Splitting B into 2^k many equal-mass atoms gives an η -sum of $2^k(k + N)/(2^k 2^N)$. Setting $k = k_N := 2^N - N$ makes this η -sum equal 1; so splitting the pieces of $X = \bigsqcup_{N=1}^{\infty} B_N$, with $\mu(B_N) = \frac{1}{2^N}$, yields an ∞ -distropy partition.

Function η

The $\eta(x) = x \log(1/x)$ function has vertical tangent at $x = 0$, maximum at $1/e$ and, when graphed in nats slope -1 at $x = 1$. (Curiosity: Just in this paragraph we compute distropy in *nats*, that is, using natural logarithm. Given a small probability $p \in [0, 1]$ and setting $x := 1/p$, note that $\eta(p) =$

$\frac{\log(x)}{x} \approx 1/\pi(x)$, where $\pi(x)$ denotes the number of prime numbers less-equal x . (This approximation is a weak form of the Prime Number Theorem.) Is there any actual connection between the “approximate distropy” function $\mathcal{H}_\pi(\vec{p}) := \sum_{p \in \vec{p}} 1/\pi(1/p)$ and Number Theory, other than a coincidence of growth rate?)

Consider partitions P and Q on the same space (X, μ) . Their **join**, written $P \vee Q$, has atoms $A \cap B$, for each pair $A \in P$ and $B \in Q$. They are **independent**, written $P \perp Q$ if $\mu(A \cap B) = \mu(A)\mu(B)$ for each A, B pair. We write $P \succcurlyeq Q$ and say that “ P refines Q ,” if each P -atom is a subset of some Q -atom. Consequently, each Q -atom is a union of P -atoms. Recall, for δ a real number, our convention that δ^c means $1 - \delta$, in analogy with $\mu(B^c)$ equaling $1 - \mu(B)$ on a probability space.

Distropy Fact

For partitions P, Q, R on probability space (X, μ) :

- (a) $\mathcal{H}(P) \leq \log(\#P)$, with equality IFF P is an equi-mass partition.

- (b) $\mathcal{H}(Q \vee R) \leq \mathcal{H}(Q) + \mathcal{H}(R)$, with equality IFF $Q \perp R$.
 (c) For $\delta \in [0, \frac{1}{2}]$, the function $\delta \mapsto \mathcal{H}(\delta, \delta^c)$ is strictly increasing.
 (d) $R \preccurlyeq P$ implies $\mathcal{H}(R) \leq \mathcal{H}(P)$, with equality IFF $R = P$.

Proof Use the strict concavity of $\eta(\cdot)$, together with Jensen’s inequality (Figs. 2 and 3).

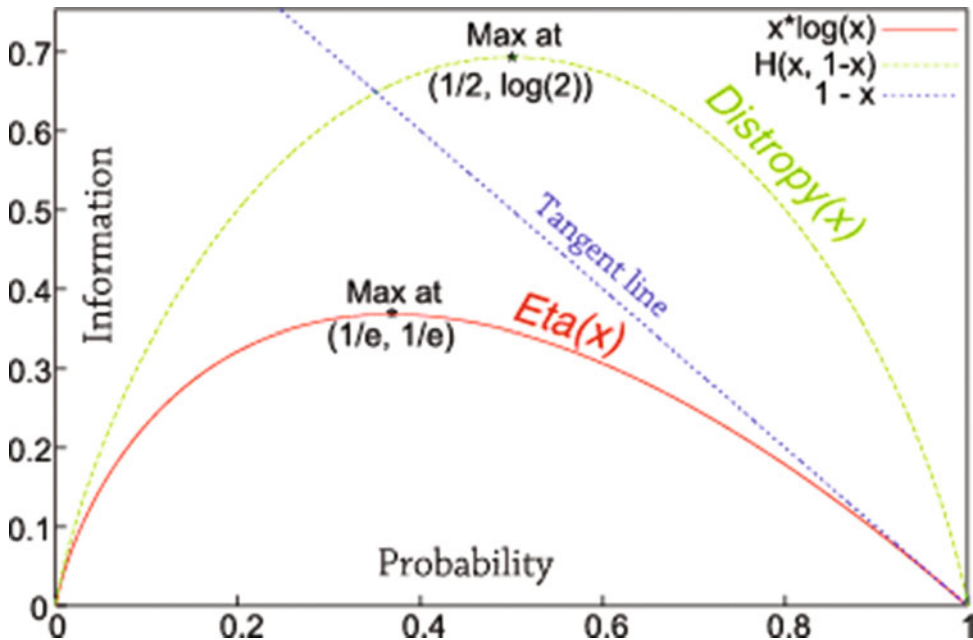
Remark 1 Although we will not discuss it in this entry, most distropy statements remain true with “partition” replaced by “countable partition of finite distropy.”

Binomial Coefficients

The dartboard gave an example where distropy arises in a natural way. Here is a second example.

For a small $\delta > 0$, one might guess that the binomial co-efficient $\binom{n}{\delta n}$ grows asymptotically (as $n \rightarrow \infty$) like $2^{\epsilon n}$, some small ϵ . But what is the correct relation between ϵ and δ ?

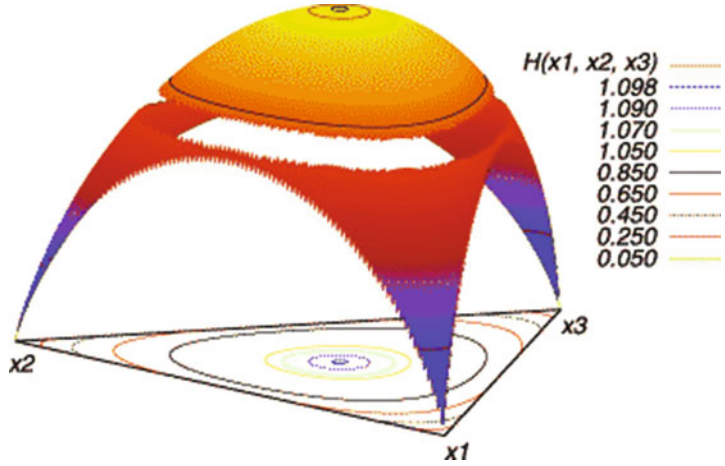
Well, Stirling’s formula $n! \approx [n/e]^n$ gives



Entropy in Ergodic Theory, Fig. 2 Using natural log, here are the graphs of: $\eta(x)$ in solid red, $\mathcal{H}(x, x^c)$ in dashed green, $1 - x$ in dotted blue. Both $\eta(x)$ and $\mathcal{H}(x, x^c)$ are strictly convex-down. The $1 - x$ line is tangent to $\eta(x)$ at $x = 1$

Entropy in Ergodic

Theory, Fig. 3 Using natural log: The graph of $\mathcal{H}(x_1, x_2, x_3)$ in barycentric coordinates; a slice has been removed, between $z = 0.745$ and $z = 0.821$. The three arches are copies of the distropy curve from Fig. 2



$$\begin{aligned} \frac{n!}{[\delta n]! [\delta^c n]!} &\approx \frac{n^n}{[\delta n]^{\delta n} [\delta^c n]^{\delta^c n}} \\ &= \frac{1}{[\delta^{\delta n} \delta^c]^{\delta^c n}} \quad (\text{recall } \delta^c = 1 - \delta). \end{aligned}$$

Thus, $\frac{1}{n} \log \binom{n}{\delta n} \approx \mathcal{H}(\delta, \delta^c)$. But by means of the above distropy inequalities, we get an inequality true for *all* n , not just asymptotically.

Lemma 2 (Binomial Lemma) Fix a $\delta \in [0, \frac{1}{2}]$ and let $\mathbf{H} := \mathcal{H}(\delta, \delta^c)$. Then for each $n \in \mathbb{Z}_+$:

$$\sum_{j \in [0, \dots, \delta n]} \binom{n}{j} \leq 2^{\mathbf{H}n}. \quad (4)$$

Proof Let $X \subset \{0, 1\}^n$ be the set of x with

$$\#\{i \in [1 \dots n] | x_i = 1\} \leq \delta n.$$

On X , let $P_1, P_2, \dots$ be the coordinate partitions; for example, $P_7 = (A_7, A_7^c)$, where $A_7 := \{x | x_7 = 1\}$. Weighting each point by $\frac{1}{|X|}$, the uniform distribution on X , gives that $\mu(A_7) \leq \delta$. So $\mathcal{H}(P_7) \leq \mathbf{H}$, by (c) in section “[Distropy Fact](#).” Finally, the join $P_1 \vee \dots \vee P_n$ separates the points of X . So

$$\begin{aligned} \log(\#X) &= \mathcal{H}(P_1 \vee \dots \vee P_n) \\ &\leq \mathcal{H}(P_1) + \dots + \mathcal{H}(P_n) \\ &\leq \mathbf{H}n, \end{aligned}$$

making use of (a), (b) in “[Distropy Fact](#).” And $\#X$ equals LhS(4).

A Glance at Shannon’s Noisy Channel Theorem

We can restate the Binomial lemma using the *Hamming metric* on $\{0, 1\}^n$,

$$\text{Dist}(x, y) := \#\{i \in [1 \dots n] | x_i \neq y_i\}$$

for all $x, y \in \{0, 1\}^n$

Use $\text{Bal}(x, r)$ for the open radius- r ball centered at x , and

$$\overline{\text{Bal}}(x, r) := \{y | \text{Dist}(x, y) \leq r\}$$

for the closed ball. The above lemma can be interpreted as saying that

$$\begin{aligned} |\overline{\text{Bal}}(x, \delta n)| &\leq 2^{\mathcal{H}(\delta, \delta^c)n}, \text{ for each } x \in \{0, 1\}^n. \end{aligned} \quad (5)$$

Corollary 3 Fix $n \in \mathbb{Z}_+$ and $\delta \in [0, \frac{1}{2}]$, and let

$$\mathbf{H} := \mathcal{H}(\delta, \delta^c).$$

Then there is a set $C \subset \{0, 1\}^n$, with $\#C \geq 2^{\lfloor 1 - \mathbf{H} \rfloor n}$, that is **strongly δn -separated**. That is, $\text{Dist}(x, y) > \delta n$ for each distinct pair $x, y \in C$.

Noisy Channel

Shannon's theorem says that a noisy channel has a **channel capacity**. Transmitting *above* this speed, there is a minimum error-rate (depending how much “above”) that no errorcorrecting code can fix. Conversely, one can transmit *below* – but arbitrarily close to – the channel capacity, and encode the data so as to make the error-rate less than any given ϵ . We use Corollary 3 to show the existence of such codes, in the simplest case where the noise is a binary independent-process (a “Bernoulli” process, in the language later in this entry). (The noise-process is assumed to be *independent* of the signal-process. In contrast, when the perturbation is highly dependent on the signal, then it is sometimes called **distortion**.)

We have a channel which can pass one bit per second. Alas, there is a fixed noise-probability $v \in [0, \frac{1}{2})$ so that a bit in the channel is perturbed into the other value. Each perturbation is independent of all others. Let $\mathbf{H} := \mathcal{H}(v, v^c)$. The value $[1 - \mathbf{H}]$ bits-per-second is the **channel capacity** of this noise-afflicted channel.

Encoding/Decoding

Encode using an “ k, n -**block-code**”; an injective map $F: \{0, 1\}^k \rightarrow \{0, 1\}^n$. The source text is split into consecutive k -bit blocks. A block $x \in \{0, 1\}^k$ is encoded to $F(x) \in \{0, 1\}^n$ and then sent through the channel, where it comes out perturbed to $\alpha \in \{0, 1\}^n$. The *transmission rate* is thus k/n bits per second.

For this example, we fix a radius $r > 0$ to determine the decoding map,

$$D_r: \{0, 1\}^n \rightarrow \{\text{Oops}\} \sqcup \{0, 1\}^k.$$

We set $D_r(\alpha)$ to z if there is a *unique* z with $F(z) \in \text{Bal}(\alpha, r)$; else, set $D_r(\alpha) := \text{Oops}$. One can

think of the noise as a $\{0, 1\}$ -independent process, with $\text{Prob}(1) = v$, which is added mod 2 to the signal-process. Suppose we can arrange that the set $\{F(x) \mid x \in \{0, 1\}^k\}$ of codewords is a strongly r -separated-set. Then

The probability that a block is mis – decoded is the probability, flipping a v – coin n times that we get more than r many Heads.

(6)

Theorem 4 (Shannon) Fix a noise-probability $v \in [0, \frac{1}{2})$ and let $\mathbf{H} := \mathcal{H}(v, v^c)$. Consider a rate $R < [1 - \mathbf{H}]$ and an $\epsilon > 0$. Then $\forall_{\text{large } n}$ there exists a k and a code $F: \{0, 1\}^k \rightarrow \{0, 1\}^n$ so that: The F -code transmits bits at faster than R bits-per-second and with error-rate $< \epsilon$.

Proof Let $\mathbf{H}' := \mathcal{H}(\delta, \delta^c)$, where $\delta > v$ was chosen so close to v that

$$\delta < \frac{1}{2} \text{ and } 1 - \mathbf{H}' > R. \quad (7)$$

Pick a large n for which

$$\frac{k}{n} > R, \text{ where } k := \lfloor [1 - \mathbf{H}']n \rfloor. \quad (8)$$

By Corollary 3, there is a strongly δn -separated set $C \subset \{0, 1\}^n$ with $\#C \geq 2^{\lfloor 1 - \mathbf{H}' \rfloor n}$. So C is big enough to permit an injection $F: \{0, 1\}^k \rightarrow C$. The probability of a decoding error is that of getting more than δn many Heads in flipping a v -coin n times (see Courtesy Eq. 6). Since $\delta > v$, the Weak Law of Large Numbers guarantees – once n is large enough – that this probability is less than the given ϵ .

The Information Function

Agree to use $\mathbf{P} = (A_1, \dots)$, $\mathbf{Q} = (B_1, \dots)$, $\mathbf{R} = (C_1, \dots)$ for partitions, and $\mathcal{F}, \mathcal{G}$ for fields.

With $\mathfrak{C}$ a (finite or infinite) family of subfields of $\mathcal{X}$, their **join** $\bigvee_{\mathcal{G} \in \mathfrak{C}} \mathcal{G}$ is the smallest field $\mathcal{F}$ such

that $\mathcal{G} \subset \mathcal{F}$, for each $\mathcal{G} \in \mathfrak{C}$. A partition Q can be interpreted also as a field namely, the field of unions of its atoms. A join of denumerably many partitions will be interpreted as a field, but a join of *finitely* many, $P_1 \vee \dots \vee P_N$, will be viewed as a partition *or* as a field, depending on context.

Conditioning a partition P on a positive-ass set B , let $P|B$ be the probability vector $A \mapsto \frac{\mu(A \cap B)}{\mu(B)}$. Its distropy is

$$\mathcal{H}(P|B) = \sum_{A \in P} \frac{\mu(A \cap B)}{\mu(B)} \log \left(\frac{1}{\mu(A \cap B)/\mu(B)} \right).$$

So conditioning P on a *partition* Q gives conditional distropy

$$\begin{aligned} \mathcal{H}(P|Q) &= \sum_{B \in Q} \mu(B) \mathcal{H}(P|B) \\ &= \sum_{A \in P, B \in Q} \mu(A \cap B) \log \left(\frac{1}{\mu(A \cap B)/\mu(B)} \right). \end{aligned} \quad (9)$$

A “dartboard” interpretation of $\mathcal{H}(P|Q)$ is

The expected number of questions to ascertain the P -atom that a random dart $x \in X$ fell in, given that we are told its Q -atom.

For a set $A \subset X$, use $\mathbf{1}_A : X \rightarrow \{0,1\}$ for its *indicator function*; $\mathbf{1}_A(x) = 1$ IFF $x \in A$. The **information function** of partition P , a map $\mathcal{I}_P : X \rightarrow [0, \infty)$, is

$$\mathcal{I}_P := \sum_{A \in P} \log \left(\frac{1}{\mu(A)} \right) \mathbf{1}_A(\cdot). \quad (10)$$

The information function has been defined so that its expectation is the distropy of P ,

$$\mathbb{E}(\mathcal{I}_P) = \int_X \mathcal{I}_P(\cdot) d\mu = \mathcal{H}(P).$$

Conditioning on a Field

For a subfield $\mathcal{F} \subset \mathcal{X}$, recall that each function $g \in \mathbb{L}^1(\mu)$ has a *conditional expectation* $\mathbb{E}(g|\mathcal{F}) \in \mathbb{L}^1(\mu)$. It is the *unique* $\mathcal{F}$ -measurable function with

$$\forall B \in \mathcal{F} : \int_B \mathbb{E}(g|\mathcal{F}) d\mu = \int_B g d\mu.$$

Returning to distropy ideas, use $\mu(A|\mathcal{F})$ for the **conditional probability** function; it is the conditional expectation $\mathbb{E}(\mathbf{1}_A|\mathcal{F})$. So the **conditional information function** is

$$\mathcal{I}_{P|\mathcal{F}}(x) := \sum_{A \in P} \log \left(\frac{1}{\mu(A|\mathcal{F})} \right) \mathbf{1}_A(x). \quad (11)$$

Its integral

$$\mathcal{H}(P|\mathcal{F}) := \int \mathcal{I}_{P|\mathcal{F}} d\mu,$$

is the **conditional distropy** of P on $\mathcal{F}$.

When $\mathcal{F}$ is the field of unions of atoms from some partition Q , then the number $\mathcal{H}(P|\mathcal{F})$ equals the $\mathcal{H}(P|Q)$ from (9).

Write $\mathcal{G}_j \nearrow \mathcal{F}$ to indicate that fields $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \dots$ are nested, and that $\text{Fld} \left(\bigcup_1^\infty \mathcal{G}_j \right) = \mathcal{F}$, a.e. The Martingale Convergence Theorem (p. 103 in (Petersen 1983)) gives (c) below.

Conditional-Distropy Fact

Consider partitions P, Q, R and fields $\mathcal{F}$ and $\mathcal{G}_j$. Then

- (a) $0 \leq \mathcal{H}(P|\mathcal{F}) \leq \mathcal{H}(P)$, with equality IFF $P \stackrel{a.e.}{\subset} \mathcal{F}$, respectively, $P \perp \mathcal{F}$. (We regard P as a field of unions of atoms for the partition P)
- (b) $\mathcal{H}(Q \vee R|\mathcal{F}) \leq \mathcal{H}(Q|\mathcal{F}) + \mathcal{H}(R|\mathcal{F})$.
- (c) Suppose $\mathcal{G}_j \nearrow \mathcal{F}$. Then $\mathcal{H}(P|\mathcal{G}_j) \searrow \mathcal{H}(P|\mathcal{F})$.
In particular, $\mathcal{H}(P|\mathcal{F}) \leq \mathcal{H}(P|\mathcal{G})$ if $\mathcal{G}$ is a subfield of $\mathcal{F}$.
- (d) $\mathcal{H}(Q \vee R) = \mathcal{H}(Q|R) + \mathcal{H}(R)$.
- (e) $\mathcal{H}(Q \vee R_1|R_0) = \mathcal{H}(Q|R_1 \vee R_0) + \mathcal{H}(R_1|R_0)$.

Imagining our dartboard (Fig. 1) divided by superimposed partitions Q and R , equality (d) can be interpreted as saying: “*You can efficiently discover where the dart landed in both partitions, by first asking efficient questions about R , then – based on where it landed in R – asking intelligent questions about Q .*”

Entropy of a Process

Consider a transformation (X, μ, T) and partition $P = (A_1, A_2, \dots)$. Each “time” n determines a partition $P_n := T^n P$, whose j th-atom is $T^{-n}(A_j)$. The **process** T, P refers to how T acts on the subfield $V_0^\infty P_n \subset X$. (An alternative view of a process is as a stationary sequence $V_0, V_1, \dots$ of random variables $V_n : X \rightarrow \mathbb{Z}_+$, where $V_n(x) := j$ because x is in the j th-atom of P_n .)

Write $h(T, P)$ for the “**entropy**” of the T, P process. It is the limit of the **conditional-distropy-numbers**

$$c_n := \mathcal{H}(P_0 | P_1 \vee P_2 \vee \dots \vee P_{n-1}).$$

This limit exists since $\mathcal{H}(P) = c_1 \geq c_2 \geq \dots \geq 0$. Define the **average-distropy-number** $\frac{1}{n}h_n$, where

$$h_n := \mathcal{H}(P_0 \vee P_1 \vee \dots \vee P_{n-1})$$

Certainly $h_n = c_n + \mathcal{H}(P_1 \vee \dots \vee P_{n-1}) = c_n + h_{n-1}$, since T is measure-preserving. Induction gives $h_n = \sum_{j=1}^n c_j$. So the Cesàro averages $\frac{1}{n}h_n$ converge to the entropy.

Theorem 5 *The entropy of process (X, P, μ, T) equals*

$$\begin{aligned} h(T, P) &:= \lim_{n \rightarrow \infty} \frac{1}{n} \mathcal{H}(P_0 \vee \dots \vee P_{n-1}) \\ &= \lim_{n \rightarrow \infty} \mathcal{H}\left(P_0 \left| \bigvee_{j=1}^n P_j \right.\right) \\ &= \mathcal{H}\left(P_0 \left| \bigvee_{j=1}^\infty P_j \right.\right). \end{aligned}$$

Both limits are nonincreasing. The entropy $h(T, P) \geq 0$, with equality IFF $P \subset V_1^\infty P$ $j \in \mathbb{G}$. And $h(T, P) \leq \mathcal{H}(P)$, with equality IFF T, P is an independent process.

Generators

We henceforth only discuss *invertible* mpts, that is, when T^{-1} is itself an mpt. Viewing the atoms of P as “letters,” then, each $x \in X$ has a T, P – **name**

$$\dots x_{-2}x_{-1}x_0x_1x_2x_3\dots,$$

where x_n is $P\langle T^n(x) \rangle$, the P -letter owning $T^n(x)$.

A partition P **generates** (the whole field) under T , if $V_{-\infty}^\infty T^n P =_\mu X$. It turns out that P generates IFF P **separates points**. The equivalence of *generating* and *separating* is a technical theorem, due to Rokhlin. Assuming μ to be Lebesgue is not much of a limitation. For instance, if μ is a finite measure on any Polish space, then μ extends to a Lebesgue measure on the μ -completion of the Borel sets. To not mince words: All spaces are Lebesgue spaces unless you are actively *looking* for trouble.) That is, after deleting a (T -invariant) nullset, distinct points of X have distinct T, P -names².

A finite set $\{1, \dots, L\}$ of integers, interpreted as an **alphabet**, yields the **shift space** $X := \{1, \dots, L\}^\mathbb{Z}$ of doubly-infinite sequences $x = (\dots x_{-1}x_0x_1\dots)$. The **shift** $T : X \rightarrow X$ acts on X by

$$T(x)_n := x_{n+1}.$$

The time-zero partition P separates points, under the action of the shift. This L -atom **time-zero partition** has $P\langle x \rangle = P\langle y \rangle$ IFF $x_0 = y_0$. So no matter what shift-invariant measure is put on X , the time-zero partition will generate under the action of T . It is Kolmogorov who formulated measure-preserving transformation $(X, \mathcal{X}, \mu, T)$ into the frame of stationary process, using a generating partition.

Time Reversibility

A transformation need not be isomorphic to its inverse. Nonetheless, the average-distropy-numbers show that $h(T^{-1}, P) = h(T, P)$, although this is not obvious from the conditioning-definition of entropy. Alternatively,

$$\begin{aligned} &\mathcal{H}\left(P_0 \left| \bigvee_{j=1}^n P_j \right.\right) \vee \\ &= \mathcal{H}(P_0 \vee \dots \vee P_n) - \mathcal{H}(P_1 \vee \dots \vee P_n) \\ &= \mathcal{H}(P_{-n} \vee \dots \vee P_0) - \mathcal{H}(P_{-n} \vee \dots \vee P_{-1}) \\ &= \mathcal{H}\left(P_0 \left| \bigvee_{j=-1}^n P_j \right.\right). \end{aligned} \tag{12}$$

²We are now at liberty to reveal that our X has always been a **Lebesgue space**, that is, measure-isomorphic to an interval, $[0, 1]$.

Bernoulli Processes

A probability vector $\vec{v} := (v_1, \dots, v_L)$ can be viewed as a measure on alphabet $\{1, \dots, L\}$. Let $\mu_{\vec{v}}$ be the resulting product measure on $X := \{1, \dots, L\}^{\mathbb{Z}}$, with T the shift on X and P the time-zero partition. The independent process (T, P) is called, by ergodic theorists, a **Bernoulli process**. It can be considered as a generalization of coin tossing. Not necessarily consistently, we tend to refer to the underlying transformation as a **Bernoulli shift**.

The $(\frac{1}{2}, \frac{1}{2})$ -Bernoulli and the $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ -Bernoulli have different process-entropies, but perhaps their underlying transformations are isomorphic? Prior to the Kolmogorov-Sinai definition of entropy of a transformation, this question remained unanswered. (This is sometimes called **measure(-theoretic)** entropy or (perhaps unfortunately) **metric entropy**, to distinguish it from topological entropy. Tools known prior to entropy, such as *spectral* properties, did *not* distinguish the two Bernoulli-shifts; see Cornfel'd et al. (1982) and ► “Spectral Theory of Dynamical Systems” for the definitions.)

Entropy of a Transformation

The Kolmogorov-Sinai definition of the entropy of an mpt is

$$h(T) := \sup_Q \{h(T, Q) \mid Q \text{ a partition on } X\}.$$

Certainly entropy is an isomorphism invariant – but is it useful? After all, the supremum of *distropies* of partitions is always infinite (on nonatomic spaces) and one might fear that the same holds for entropies. The key observation (restated in Lemma 8c and proved below) was this, from (Kolmogorov 1958; Sinai 1959).

Theorem 6 (Kolmogorov-Sinai Theorem) *If P generates under T , then $h(T) = h(T, P)$.*

Thereupon the $(\frac{1}{2}, \frac{1}{2})$ and $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ Bernoulli-shifts are *not* isomorphic, since their respective entropies are $\log(2) \neq \log(3)$.

Wolfgang Krieger later proved a converse to the Kolmogorov-Sinai theorem (Krieger 1970).

Theorem 7 (Krieger Generator Theorem, 1970) *Suppose T ergodic. If $h(T) < \infty$, then T has a finite generating partition. Indeed, letting K be the smallest integer $K > h(T)$, there is a K -atom generator. (It is an easier result, undoubtedly known much earlier, that every ergodic T has a countable generating partition – possibly of ∞ -distropy.)*

Proof See § 7.5 in Rudolph (1990), where Krieger’s theorem is stated in terms of joinings.

Entropy Is Continuous

Given ordered partitions $Q = (B_1, \dots)$ and $Q' = (B'_1, \dots)$, extend the shorter by null-atoms until $|Q| = |Q'|$. Let $\text{Fat} := \bigcup_j B_j \cap B'_j$; this set should have mass close to 1 if Q and Q' are almost the same partition. Define a new partition

$$Q \nabla Q' := \{\text{Fat}\} \sqcup \left\{ B_i \cap B'_j \mid \text{with } i \neq j \right\}.$$

(In other words, take $Q \vee Q'$ and coalesce, into a single atom, all the $B_k \cap B'_k$ sets.) Topologize the space of partitions by saying that $Q^{(L)} \rightarrow Q$ when $\mathcal{H}(Q \nabla Q^{(L)}) \rightarrow 0$. (On the set of ordered K -set partitions (with K fixed) this convergence is the same as: $Q^{(L)} \rightarrow Q$ when $\mu(\text{Fat}(Q^{(L)}, Q)) \rightarrow 1$. An alternative approach is the **Rokhlin metric**, $\text{Dist}(P, Q) := \mathcal{H}(P \mid Q) + \mathcal{H}(Q \mid P)$, which has the advantage of working for *unordered* partitions.) Then Lemma 8b says that process-entropy varies continuously with varying the partition.

Lemma 8 *Fix a mpt (X, μ, T) . For partitions P, Q, Q' , define $R := Q \nabla Q'$ and let $\delta := \mathcal{H}(R)$. Then*

- (a) $|\mathcal{H}(Q) - \mathcal{H}(Q')| \leq \delta$. (*Distropy varies continuously with the partition.*)
- (b) $|h(T, Q) - h(T, Q')| \leq \delta$. (*Process-entropy varies continuously with the partition.*)
- (c) *For all partitions $Q \subset \bigvee_{-\infty}^{\infty} T^{-i}P$: $h(T, Q) \leq h(T, P)$.*

In particular, $h(T, V_{-L}^L P_L) = h(T, P)$.

Proof(a) Evidently $Q' \vee R = Q' \vee Q = Q \vee R$. So
 $\mathcal{H}(Q') \leq \mathcal{H}(Q \vee R) \leq \mathcal{H}(Q) + \delta$
 (b) As above,

$$\mathcal{H}\left(\bigvee_1^N Q_j\right) \leq \mathcal{H}\left(\bigvee_1^N Q_j\right) + \mathcal{H}\left(\bigvee_1^N R_j\right).$$

Sending $N \rightarrow \infty$ gives $h(T, Q') \leq h(T, Q) + h(T, R)$. Finally, $h(T, R) \leq \mathcal{H}(R)$ and so $h(T, Q') \leq h(T, Q) + \delta$.

(c) Let $K := |Q|$. Then there is a sequence of K -set partitions $Q^{(L)} \rightarrow Q$ with $Q^{(L)} \leq \bigvee_{-L}^L P_\ell$. By above, $h(T, Q^{(L)}) \rightarrow h(T, Q)$, so showing that

$$h\left(T, \bigvee_{-L}^L P_\ell\right) \stackrel{?}{\leq} h(T, P)$$

will suffice. Note that

$$h_N := \mathcal{H}\left(\bigvee_{n=0}^{N-1} T^n \left(\bigvee_{-L}^L P_\ell\right)\right) = \mathcal{H}\left(\bigvee_{j=-L}^{N-1+L} P_j\right).$$

So

$$\frac{1}{N} h_N \leq \frac{1}{N} \mathcal{H}(V_0^{N-1} P_j) + \frac{1}{N} 2L \mathcal{H}(P).$$

Now send $N \rightarrow \infty$.

Entropy Is Not Continuous

The most common topology placed on the space Ω of mpts is the *coarse topology* that Halmos discusses in his “little red book” (Halmos 1956). (i.e., $S_n \rightarrow T$ IFF $\forall A \in \mathcal{X} : \mu(S_n^{-1}(A) \Delta T^{-1}(A)) \rightarrow 0$; this is a metric-topology, since our probability space is countably generated. This can be restated in terms of the unitary operator U_T on $\mathbb{L}^2(\mu)$, where $U_T(f) := f \circ T$. Namely, $S_n \rightarrow T$ in the coarse topology IFF $U_{S_n} \rightarrow U_T$ in the strong operator topology.)

The Rokhlin lemma (see p. 33 in Petersen 1983) implies that the isomorphism-class of *each* ergodic mpt is *dense* in Ω , (e.g., see p. 77 in Halmos 1956) disclosing that the $S \mapsto h(S)$ map is discontinuous.

Indeed, the failure happens already for process-entropy with respect to a fixed partition. A Bernoulli process T, P has positive entropy. Take mpts $S_n \rightarrow T$, each isomorphic to an irrational rotation. Then each $h(S_n, P)$ is zero, as shown in the later section “[Determinism and Zero-Entropy](#).”

Further Results

When $\mathcal{F}$ is a T invariant subfield, agree to use “ $T|_{\mathcal{F}}$ for T -restricted to $\mathcal{F}$,” which is a factor (see section “[Glossary](#)”) of T . Transformations T and S are **weakly isomorphic** if each is isomorphic to a factor of the other.

The foregoing entropy tools make short shrift of the following.

Lemma 9 (Entropy Lemma) Consider T -invariant sub-fields $\mathcal{G}_j$ and $\mathcal{F}$.

(a) Suppose $\mathcal{G}_j \nearrow \mathcal{F}$. Then $h(T|_{\mathcal{G}_j}) \nearrow h(T|_{\mathcal{F}})$. In particular, $\mathcal{G} \subset \mathcal{F}$ implies that $h(T|_{\mathcal{G}}) \leq h(T|_{\mathcal{F}})$, so entropy is an invariant of weak-isomorphism.

(b) $h(T|_{\mathcal{G}_1 \vee \mathcal{G}_2 \vee \dots}) \leq \sum_j h(T|_{\mathcal{G}_j})$.

And $h(T, Q_1 \vee Q_2 \vee \dots) \leq \sum_j h(T, Q_j)$.

(c) For mpts $(Y_j, \nu_j, S_j) : h(S_1 \times S_2 \times \dots) = \sum_j h(S_j)$.

(d) $h(T^{-1}) = h(T)$. More generally, $h(T^n) = |n| \cdot h(T)$.

Meshalkin’s Map

In the wake of Kolmogorov’s 1958 entropy paper, for two Bernoulli-shifts to be isomorphic one now knew that they had to have equal entropies. Meshalkin provided the first nontrivial example in 1959 (Meshalkin 1959).

Let $S : Y \rightarrow Y$ be the Bernoulli-shift over the “letter” alphabet $\{E, D, P, N\}$, with probability distribution $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. The letters E, D, P, N stand for *Even, Odd, Positive, Negative*, and will be used to describe the code (isomorphism) between the processes.

Use $T : X \rightarrow X$ for the Bernoulli-shift over “digit” alphabet $\{0, +1, -1, +2, -2\}$, with probability distribution $(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$. Both distributions $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and $(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$ have distropy $\log(4)$.

We denote S by $B(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and T by $B(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$. After deleting invariant nullsets from X and Y , the construction will produce a measure-preserving isomorphism $\psi : X \rightarrow Y$ so that $T \circ \psi = \psi \circ S$.

The Code

In X , consider this point x :

... 0 0 0 - 1 0 0 + 1 + 2 - 1 + 1 0 ...

Regard each 0 as a left-parenthesis, and each nonzero as a right-parenthesis. Link them according to the legal way of matching parentheses, as shown in the top row, below:

0	0	0	-	1	0	0	+	1	+	2	-	1	+	1	0	?
P	N	N		D	P	P	D	E		D	D					

The leftmost 0 is linked to the rightmost +1, as indicated by the longest-overbar. The left/right-parentheses form a $(\frac{1}{2}, \frac{1}{2})$ -random-walk. Since this random walk is recurrent, every position in x will be linked (except for a nullset of points x).

Below each 0, write “P” or “N” as the 0 is linked to a *positive* or *negative* digit. And below the other digits, write “E” or “D” as the digit is *even* or *odd*. So the upper name in X is mapped to the lower name, a point $y \in Y$.

This map $\psi : X \rightarrow Y$ carries the upstairs $B(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$ to $B(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, downstairs. It takes some arguing to show that independence is preserved.

The inverse map ψ^{-1} views D and E as right-parentheses, and P and N as left. Above D , write

the odd digit +1 or -1, as this D is linked to *Positive* or *Negative*.

Markov Shifts

A Bernoulli process T, P has independence $P_{(-\infty \dots 0]} \perp P_1$, whereas a **Markov process** is a bit less aloof:

The infinite Past $P_{(-\infty \dots 0]}$ doesn't provide any more information about Tomorrow than Today did.

That is, the conditional distribution $P_1 \mid P_{(-\infty \dots 0]}$ equals $P_1 \mid P_0$. Equivalently,

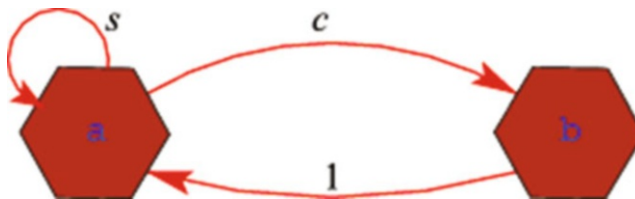
$$\mathcal{H}(P_1|P_0) = \mathcal{H}(P_1|P_{(-\infty \dots 0]}) \stackrel{\text{note}}{=} h(T, P). \quad (13)$$

The simplest nontrivial Markov process (X, P, μ, T) is over a two-letter alphabet $\{a, b\}$ and has transition graph Fig. 4, for some choice of transition probabilities s and c . The graph's Markov matrix is

$$M = [m_{ij}]_{i,j} = \begin{Bmatrix} s & c \\ 1 & 0 \end{Bmatrix},$$

where $s + c = 1$, and $m_{i,j}$ denotes the probability of going from state i to state j .

If Today's distribution on the two states is the probability-vector $\vec{v} := [p_a \ p_b]$, then Tomorrow's is the product $\vec{v}M$. So a *stationary* process needs $\vec{v}M = \vec{v}$. This equation has the unique solution $p_a = \frac{1}{1+c}$ and $p_b = \frac{c}{1+c}$. An example of computing the probability of a word or cylinder set (see section “The Caratheodory Construction” in “Measure-Preserving Systems”) in the process, is



Entropy in Ergodic Theory, Fig. 4 Call the transition probabilities $s := \text{Prob}(a \rightarrow a)$ for stay, and $c := \text{Prob}(a \rightarrow b)$ for change. These are nonnegative reals, and $s + c = 1$

$$\begin{aligned}\mu_s(\text{baaaba}) &= p_b m_{ba} m_{aa} m_{ab} m_{ba} \\ &= \frac{c}{1+c} \cdot 1 \cdot s \cdot s \cdot c \cdot 1.\end{aligned}$$

The subscript on μ_s indicates the dependence on the transition probabilities; let's also mark the mpt and call it T_s . Using (13), the entropy of our Markov map is

$$\begin{aligned}h(T_s) &= p_a \cdot \mathcal{H}(s, c) + p_b \cdot \underbrace{\mathcal{H}(1, 0)}_{=0} \\ &= \frac{-1}{1+c} \cdot [s \log(s) + c \log(c)].\end{aligned}\quad (14)$$

Determinism and Zero-Entropy

Irrational rotations have zero-entropy; let's reveal this in two different ways.

Equip $X := [0, 1)$ with “length” (Lebesgue) measure and wrap it into a circle. With “ $\oplus$ ” denoting addition mod 1, have $T: X \rightarrow X$ be the rotation $T(x) := x \oplus \alpha$, where the rotation number α is irrational. Pick distinct points $y_0, z_0 \in X$, and let P be the partition whose two atoms are the intervals $[y_0, z_0]$ and $[z_0, y_0]$, wrapping around the circle.

The T -orbit of each point x is dense in X . (Fix an $\varepsilon > 0$ and an $N > 1/\varepsilon$. Points $x, T(x), \dots, T^N(x)$ have some two at distance less than $\frac{1}{N}$; say, $\text{Dist}(T^i(x), T^j(x)) < \varepsilon$, for some $0 \leq i < j \leq N$. Since T is an isometry, $\varepsilon > \text{Dist}(x, T^k(x)) > 0$, where $k := j - i$. So the T^k -orbit of x is ε -dense.) In particular, y_0 has dense orbit, so P separates points-hence generates-under T . Our goal, thus, is $h(T, P) \stackrel{?}{=} 0$.

Rotations Are Deterministic

The forward T -orbit of each point is dense and so is the backward T -orbit. That is, the backward T , P -name of each x actually tells us which point x is. That is, $P \subset V_{-\infty}^{-1} T^n P$, which is our definition of “process T, P is **deterministic**.” Our P being finite, this determinism implies that $h(T, P)$ is zero, by Theorem 5.

Counting Names in a Rotation

The $P_0 \vee \dots \vee P_{n-1}$ partition places n translates of points y_0 and of z_0 , cutting the circle into at most

$2n$ intervals. Thus, $\mathcal{H}(P_0 \vee \dots \vee P_{n-1}) \leq \log(2n)$. And $\frac{1}{n} \log(2n) \rightarrow 0$.

Alternatively, the Shannon-McMillan-Breiman Theorem implies, for an ergodic process T, P , that the number of length- n names is approximately $2^{h(T,P)n}$; this, after discarding small mass from the space. But the growth of $n \mapsto 2n$ is linear and so, for our rotation, $h(T, P)$ must be zero.

Theorem 10 (Shannon-McMillan-Breiman Theorem (SMB-Theorem)) *Set $h := h(T, P)$, where (X, P, μ, T) is an ergodic process. Then the average information function*

$$\frac{1}{n} \mathcal{I}_{P_{[0..n)}}(x) \xrightarrow{n \rightarrow \infty} h, \text{ for a.e. } x \in X. \quad (15)$$

The functions $f_n := \mathcal{I}_{P_{[0..n)}}(x)$ converge to the constant function h both in the $\mathbb{L}^1$ -norm and in probability. (In engineering circles, this is called the almost-everywhere equipartition theorem.)

Proof See the texts of Karl Petersen (p. 261 in Petersen 1983), or Dan Rudolph (p. 77 in Rudolph 1990). See also Paul C. Shields (1996).

Consequences

Recall that $P_{[0..n)}$ means $P_0 \vee P_1 \vee \dots \vee P_{n-1}$, where $P_j := T^j P$. As usual, $P_{[0..n)}(x)$ denotes the $P_{[0..n)}$ -atom owning x .

Having deleted a nullset, we can restate (15) to now say that $\forall \varepsilon, \forall x, \forall \text{large } n$:

$$2^{-(h+\varepsilon)n} \leq \mu(P_{[0..n)}(x)) \leq 2^{-(h-\varepsilon)n}. \quad (16)$$

This has the following consequence. Fixing a number $\delta > 0$, we consider any set with $\mu(B) \geq \delta$ and count the number of n -names of points in B . The SMB-Thm implies

$$\begin{aligned}\forall \varepsilon, \forall \text{large } n, \forall B \stackrel{\mu}{\geq} \delta : \\ |\{n\text{-names in } B\}| \\ \geq 2^{(h-\varepsilon)n}.\end{aligned}\quad (17)$$

Rank-1 Has Zero-Entropy

There are several equivalent definitions for “rank-1 transformation,” several of which are discussed in

the introduction of King (1988). (See Chap. 6 in Friedman (1970) as well as Ferenczi (1997) and Shields (1973) for examples of stacking constructions.)

A **rank-1 transformation** (X, μ, T) admits a generating partition P and a sequence of Rokhlin stacks $S_n \subset X$, with heights going to ∞ , and with $\mu(S_n) \rightarrow 1$. Moreover, each of these Rokhlin stacks is P -monochromatic, that is, each level of the stack lies entirely in some atom of P .

Taking a stack of some height $2n$, let $B = B_n$ be the union of the bottom n levels of the stack. There are at most n many length- n names starting in B_n , by monochromaticity. Finally, $\mu(B_n)$ is almost $\frac{1}{2}$, so is certainly larger than $\delta := \frac{1}{3}$. Thus, Eq. (17) shows that our rank-1 T has zero entropy.

Cautions on Determinism's Relation to Zero-Entropy

A finite-valued process T, P has zero-entropy iff $P \subset V_{-\infty}^{-1} P_j$. Iterating gives

$$\bigvee_0^{\infty} P_j \subset \bigvee_{-\infty}^{-1} P_j,$$

that is, the future is measurable with respect to the past.

This was the case with the rotation, where a point's past uniquely identified the point, thus telling us its future.

While determinism and zero-entropy mean the same thing for finite-valued processes, this fails catastrophically for real-valued (i.e., continuum-valued) processes, as shown by an example of one of the authors. A stationary real-valued process $V = \dots V_{-1} V_0 V_1 V_2 \dots$ is constructed in King (1992) which is simultaneously

strongly deterministic: The two values V_0, V_1 determine all of V , future and past.

and **nonconsecutively independent**. This latter means that for each bi-infinite increasing integer sequence $\{n_j\}_{j=-\infty}^{\infty}$ with no consecutive pair (always $1 + n_j < n_{j+1}$), then the list of random variables $\dots V_{n-1} V_{n_0} V_{n_1} V_{n_2} \dots$ is an independent process.

Restricting the random variables to be *countably-valued*, how much of the example survives? Joint work with Kalikow (Kalikow and King 1994) produced a countably-valued stationary V which is nonconsecutively independent as well as deterministic. (Strong determinism is ruled out, due to cardinality considerations.) A side-effect of the construction is that V 's time-reversal $n \mapsto V_{-n}$ is **not** deterministic.

The Pinsker-Field and K-Automorphisms

Consider the collection of **zero-entropy sets**,

$$Z = Z_T := \{A \in \mathcal{X} \mid h(T, (A, A^c)) = 0\}. \quad (18)$$

Courtesy of Lemma 9b, Z is a T -invariant field, and

$$\forall Q \subset Z : h(T, Q) = 0. \quad (19)$$

The **Pinsker field** of T is this Z . It is maximal with respect to (19). (Traditionally, this called the *Pinsker algebra* where, in this context, "algebra" is understood to mean " σ -algebra.") Unsurprisingly, the **Pinsker factor** $T|_Z$ has zero entropy, that is, $h(T|_Z) = 0$. A transformation T is said to have **completely positive entropy** if it has no (nontrivial) zero-entropy factors. That is, its Pinsker field Z_T is the trivial field, $\emptyset := \{\emptyset, X\}$.

K-Processes

Kolmogorov introduced the notion of a **K-process** or **Kolmogorov process**, in which the present becomes asymptotically independent of the distant past. The asymptotic past of the T, P process is called its **tail field**, where

$$\text{Tail}(T, P) := \bigcap_{M=1}^{\infty} \bigvee_{j=-\infty}^{-M} P_j.$$

This T, P is a K-process if $\text{Tail}(T, P) = \emptyset$. This turns out to be equivalent to what we might call a strong form of "*sensitive dependence on*

initial conditions”: For each fixed length L , the distant future

$$\bigvee_{j \in (G \dots G+L]} P_j$$

becomes more and more independent of

$$\bigvee_{j \in (-\infty \dots 0]} P_j,$$

as the gap $G \rightarrow \infty$.

A transformation T is a **Kolmogorov automorphism** if it possesses a generating partition P for which T, P is a K-process.

Here is a theorem that relates the “asymptotic forgetfulness” of $\text{Tail}(T, P) = \emptyset$, to the lack of determinism implied by having no zero-entropy factors (see Walters (1981), p. 113. Related results appear in Berg (1975)).

Theorem 11 (Pinsker-Algebra Theorem) *Suppose P is a generating partition for an ergodic T . Then $\text{Tail}(T, P)$ equals Z_T .*

Since Z_T does not depend on P , this means that all generating partitions have the same tail field, and therefore K-ness of T can be detected from any generator.

Another nonevident fact follows from the above. The **future field** of T, P is defined to be $\text{Tail}(T^{-1}, P)$. It is not obvious that if the present is independent of the distant past, then it is automatically independent of the distant future. (Indeed, the precise definitions are important; witness section “[Cautions on Determinism’s Relation to Zero-Entropy](#).”) But since the entropy of a process, $h(T, (A, A^c))$, equals the entropy of the time-reversed process $h(T^{-1}, (A, A^c))$, it follows that Z_T equals $Z_{T^{-1}}$.

Ornstein Theory

In 1970, Don Ornstein (1970) solved the long-standing problem of showing that entropy was a

complete isomorphism-invariant of Bernoulli transformations; that is, that two independent processes with same entropy necessarily have the same underlying transformation. Earlier, Sinai (1962) had shown that two such Bernoulli maps were **weakly isomorphic**, that is, each isomorphic to a factor of the other.

Ornstein introduced the notion of a process being **finitely determined**, see Ornstein (1970) for a definition, proved that a transformation T was Bernoulli *IFF* it had a finitely determined generator *IFF* every partition was finitely determined with respect to T and showed that entropy completely classified the finitely determined processes up to isomorphism.

Later he developed the equivalent property, “very weak Bernoulli,” which is much easier to check and hence used often to prove the Bernoullicity (Ornstein and Weiss 1974). Moreover, the definition distinguishes Bernoullicity from Kolmogorov automorphisms. To define these properties, he introduced $\bar{d}$ -metric between two processes, (X, P, μ, T) and (Y, Q, ν, S) , where $|P| = |Q|$. If two processes are $\bar{d}$ -close, then there exists $\phi : X \rightarrow Y$ such that for all large n , the average Hamming metric between n -names of $\phi(P)$ and Q are close except on a set of small measure.

This seminal result and the ideas in the proof led to a vast machinery for proving transformations to be Bernoulli, as well as classification and structure theorems (Ornstein 1974; Shields 1973; Thouvenot 1975). Showing that the class of K-automorphisms far exceeds the Bernoulli maps, Ornstein and Shields produced in Ornstein and Shields (1973) an uncountable family of nonisomorphic K-automorphisms all with the same entropy.

Skew Product and Random Walk in Random Scenery

Let $(X, \mathcal{X}, \mu, T)$ and $(Y, \mathcal{Y}, \nu, S)$ be two measure-preserving dynamical systems. We call $\hat{T}$ on $(\hat{X}, \hat{\mathcal{X}}, \hat{\mu}) = (X \times Y, \mathcal{X} \times \mathcal{Y}, \mu \times \nu)$ a skew product if it is defined as follows

$$\widehat{T}(x, y) = (Tx, S^{\phi(x)}y)$$

where ϕ is a measurable map from $X \rightarrow \mathbb{Z}$.

The following is easy to prove

- (a) If $\phi(x) \equiv 1$ for a.e. x , then $\widehat{T} = T \times S$ and $h(\widehat{T}) = h(T) + h(S)$.
- (b) If $\int \phi(x) d\mu = 0$, then $h(\widehat{T}) = h(T)$.
- (c) If $\int \phi(x) d\mu = \infty$, then $h(\widehat{T}) = \infty$.

We mention that this is a special kind of skew product coming from two measure-preserving systems. In general, if $(\tilde{X}, \tilde{\mathcal{X}}, \tilde{\mu}, \tilde{T}) \xrightarrow{\phi} \times (X, \mathcal{X}, \mu, T)$ is a factor map, then there exists Y such that $\tilde{T}$ on $\tilde{X}$ can be represented as a skew product on $X \times Y$:

$$\tilde{T}(x, y) = (Tx, S_x y)$$

where S_x is a map on $\phi^{-1}(x)$. (We have not discussed the measure $\tilde{\mu}_x$ on the fiber $\phi^{-1}(x)$.)

Ornstein and Shields have shown uncountably many nonisomorphic K -processes of same entropy which are not Bernoulli. Their examples are via cutting and stacking method which aims at the specific properties. We let both $X = Y = \{-1, 1\}$ with $\{\frac{1}{2}, \frac{1}{2}\}$ independent probability distribution. We define on $\widehat{T} = (\widehat{X}, \widehat{P}, \widehat{\mu}) = (X \times Y, P \times Q, \mu \times \mu)$ as follows

$$\widehat{T}(x, y) = \begin{cases} (Tx, Sy) & \text{if } x_0 = 1 \\ (Tx, S^{-1}y) & \text{if } x_0 = -1 \end{cases}$$

This is called $\{T, T^{-1}\}$ -transformation. This is a well-known “natural” example of a K -automorphism, but it took some time to prove that it is not Bernoulli (Kalikow 1982).

Theorem 12 *The $\{T, T^{-1}\}$ -process is not Bernoulli.*

Main idea of the proof is that $\{T, T^{-1}\}$ -transformation does not have the property of

very weak Bernoulli which is equivalent to the property of finitely determined.

We note that $\mathcal{H}\left(\bigvee_{i=0}^{n-1} \widehat{T}^i \widehat{P} \mid \bigvee_{i=0}^{n-1} T^{-i} P\right)$ is in the order of $\sqrt{n}$; hence, additional scenery contribution to the entropy is 0. That is, $h(\widehat{\widehat{T}}) = h(T)$. After this first nonconstructive example, many smooth examples are shown to be K -automorphisms, but not Bernoulli (Katok 1980). Recently random walks in $\mathbb{Z}^2$ with $\mathbb{Z}^2$ -sceneries were investigated in den Hollander et al. (2006) and shown to have the property of K -automorphism but not Bernoulli.

When we generalize a scenery space to an independent process (not necessarily $B(\frac{1}{2}, \frac{1}{2})$ independent process), we call the process a random walk with random scenery (RWRS). They are all known to be K of same entropy. However, until recently it was not known if they are isomorphic. We have the following theorem by T. Austin (2014).

Theorem 13 *If (X, P, μ, T) is an $\{\frac{1}{2}, \frac{1}{2}\}$ independent process and (Y_1, Q_1, ν_1, S_1) and (Y_2, Q_2, ν_2, S_2) are independent process, then $\widehat{T}_1$ on $X \times Y_1$ and $\widehat{T}_2$ on $X \times Y_2$ defined in are isomorphic if and only if $\mathcal{H}(Q_1) = \mathcal{H}(Q_2)$.*

This is a finer result than Ornstein and Shields in the sense that the uncountable family of non-isomorphic K -automorphisms are the mixing extension of a K -automorphism (in fact, random walk space). He proves the above theorem by introducing a new isomorphic invariant. Instead of explaining the invariant which is quite technical, we will describe two main ingredients in the argument. The first is the mutual (shared) information defined as

$$\mathcal{I}(T, \widehat{P}) = \mathcal{H}\left(\bigvee_{i=-1}^{-n} \widehat{T}^{-i} \widehat{P}\right) + \mathcal{H}\left(\bigvee_{i=0}^{n-1} \widehat{T}^{-i} \widehat{P}\right) - \mathcal{H}\left(\bigvee_{i=-n}^n \widehat{T}^{-i} \widehat{P}\right)$$

of the n -step past and n -step future of $\widehat{P}$. Since (T, P) is an independent process,

$$\begin{aligned} & \mathcal{H}\left(\bigvee_{i=-1}^{-n} T^{-i}\mathbf{P}\right) + \mathcal{H}\left(\bigvee_{i=0}^{n-1} T^{-i}\mathbf{P}\right) \\ & - \mathcal{H}\left(\bigvee_{i=-n}^n T^{-i}\mathbf{P}\right) \\ & = 0. \end{aligned}$$

Hence, the mutual information comes from only the scenery part. This makes it possible for us to approximate (recognize) the scenery entropy if n is sufficiently large.

The second is that for sufficiently small number δ , the number of δ -balls in $\overline{d}$, $B_{\delta,n}(A) = \{B|\overline{d}(A_0^{n-1}, B_0^{n-1}) < \delta, B \in \bigvee_{i=0}^{n-1} T^{-i}\mathbf{P}\}$ also gives nice to almost the entropy. (The second is the generalization of Shannon-McMillan Theorem.)

Topological Entropy

Adler, Konheim, and McAndrew, in 1965, published the first definition of *topological entropy* in the eponymous article (Adler et al. 1965). Here, $T: X \rightarrow X$ is a continuous self-map of a compact topological space. The role of atoms is played by open sets. Instead of a finite partition, one uses a finite **open-cover** $\mathcal{V} = \{U_j\}_{j=1}^L$, that is, each **patch** U_j is open, and their union $\cup(\mathcal{V}) = X$. (Henceforth, “cover” means “open cover.”) (Because only on a compact space, we can omit “finite.” Some generalizations of topological entropy to noncompact spaces require that only *finite* open-covers be used (Hasselblatt et al. 2005).)

Two continuous maps $T: X \rightarrow X$ and $S: Y \rightarrow Y$ are **topologically conjugate** (as *isomorphism* is called in this category) if there exists a homeomorphism $\psi: X \rightarrow Y$ with $\psi T = S\psi$. We say Y is a factor of X if ψ is an onto homomorphism.

Let $\text{Card}(\mathcal{V})$ be the *minimum* cardinality over all subcovers.

$$\text{Card}(\mathcal{V}) := \text{Min}\{\#\mathcal{V}' | \mathcal{V}' \subset \mathcal{V} \text{ and } \cup(\mathcal{V}') = X\},$$

and let

$$\mathcal{H}(\mathcal{V}) = \mathcal{H}_{\text{top}}(\mathcal{V}) := \log(\text{Card}(\mathcal{V})).$$

Analogous to the definitions for partitions, define

$$\mathcal{V} \vee \mathcal{W} := \{V \cap W | V \in \mathcal{V} \text{ and } W \in \mathcal{W}\};$$

$$T\mathcal{V} := \{T^{-1}(U) | U \in \mathcal{V}\}$$

and $\mathcal{V}_{[0..n]} := \mathcal{V}_0 \vee \mathcal{V}_1 \vee \cdots \vee \mathcal{V}_{n-1}$ where $\mathcal{V}_i = T^i \mathcal{V}_0$; $\mathcal{W} \succcurlyeq \mathcal{V}$, if each $\mathcal{W}$ -patch is a subset of some $\mathcal{V}$ -patch.

The T , $\mathcal{V}$ **entropy** is

$$\begin{aligned} h(T, \mathcal{V}) &= h_{\text{top}}(T, \mathcal{V}) \\ &:= \limsup_{n \rightarrow \infty} \frac{1}{n} \mathcal{H}_{\text{top}}(\mathcal{V}_{[0..n]}). \end{aligned} \quad (20)$$

And the **topological entropy** of T is

$$h_{\text{top}}(T) := \sup_{\mathcal{V}} h_{\text{top}}(T, \mathcal{V}), \quad (21)$$

taken over all open covers $\mathcal{V}$.

Thus, h_{top} counts, in some sense, the growth rate or the divergence rate in the T -orbits of length n as in the case of measure space. Evidently, topological entropy is an isomorphism invariant (Adler and Marcus 1979; Adler and Weiss 1967).

Lemma 14 (Subadditive Lemma) *Consider a sequence $\mathbf{s} = (s_l)_1^\infty \subset [-\infty, \infty]$ satisfying $s_{k+l} \leq s_k + s_l$, for all $k, l \in \mathbb{Z}$. Then the following limit exists in $[-\infty, \infty]$, and $\lim_{n \rightarrow \infty} \frac{s_n}{n} = \inf_n \frac{s_n}{n}$.*

Topological entropy, or “**top-ent**” for short, satisfies many of the relations of measure-entropy.

Lemma 15(a) $\mathcal{V} \preccurlyeq \mathcal{W}$ implies

$$\mathcal{H}(\mathcal{V}) \leq \mathcal{H}(\mathcal{W}) \text{ and } h(T, \mathcal{V}) \leq h(T, \mathcal{W}).$$

$$(b) \mathcal{H}(\mathcal{V} \vee \mathcal{W}) \leq \mathcal{H}(\mathcal{V}) + \mathcal{H}(\mathcal{W}).$$

$$(c) \mathcal{H}(T(\mathcal{V})) \leq \mathcal{H}(\mathcal{V}), \text{ with equality if } T \text{ is surjective. Also, } h(T, \mathcal{V}) \leq \mathcal{H}(\mathcal{V}).$$

$$(d) \text{ In (20), the } \lim_{n \rightarrow \infty} \frac{1}{n} \mathcal{H}(\mathcal{V}_{[0..n]}) \text{ exists.}$$

(e) Suppose T is a homeomorphism. Then

$$h(T^{-1}, \mathcal{V}) = h(T, \mathcal{V}),$$

for each cover $\mathcal{V}$. Consequently, $h_{\text{top}}(T^{-1}) = h_{\text{top}}(T)$.

(f) Suppose $\mathfrak{E}$ is a collection of covers such that: For each cover $\mathcal{W}$, there exists a $\mathcal{V} \in \mathfrak{E}$ with $\mathcal{V} \succ \mathcal{W}$. Then $h_{\text{top}}(T)$ equals the supremum of $h_{\text{top}}(T, \mathcal{V})$, just taken over those $\mathcal{V} \in \mathfrak{E}$.

(g) For all $\ell \in \mathbb{N}$:

$$h_{\text{top}}(T^\ell) = \ell h_{\text{top}}(T)$$

Proof (c) Let $C \leq \mathcal{V}$ be a min-cardinality subcover. Then TC is a subcover of $T\mathcal{V}$. So $\text{Card } T\mathcal{V} \leq |\text{Card } TC| = |C|$. As for entropy, inequality (b) and the foregoing give $\mathcal{H}(\mathcal{V}_{[0\dots n]}) \leq n\mathcal{H}(\mathcal{V})$.

(d) Set $s_n := \mathcal{H}(\mathcal{V}_{[0\dots n]})$. Then

$$s_{k+l} \leq s_k + \mathcal{H}(T^k(\mathcal{V}_{[0\dots l]})) \leq s_k + s_l,$$

by (b) and (c), and so the Subadditive Lemma 14, applies. (g) WLOG, $\ell = 3$. Given $\mathcal{V}$ a cover, triple it to $\widehat{\mathcal{V}} := \mathcal{V} \cap T\mathcal{V} \cap T^2\mathcal{V}$; so

$$\bigvee_{j \in [0\dots N]} [T^3]^j(\widehat{\mathcal{V}}) = \bigvee_{i \in [0\dots 3N]} T^i(\mathcal{V}).$$

Thus, $\mathcal{H}(T^3, \widehat{\mathcal{V}}, N) = \mathcal{H}(T, \mathcal{V}, 3N)$, extending notation. Part (d) and sending $N \rightarrow \infty$, gives $h(T^3, \widehat{\mathcal{V}}) = 3h(T, \mathcal{V})$. Lastly, take covers such that

$$\begin{aligned} h(T^3, C^{(k)}) &\rightarrow h_{\text{top}}(T^3) \quad \text{and} \\ h(T, D^{(k)}) &\rightarrow h_{\text{top}}(T), \end{aligned}$$

as $k \rightarrow \infty$. Define $\mathcal{V}^{(k)} := C^{(k)} \vee D^{(k)}$. Apply the above to $\mathcal{V}^{(k)}$, then send $k \rightarrow \infty$.

Using a Metric

From now on, our space is a compact metric space (X, d) .

Dinaburg (1970) and Bowen (1971, 1973) gave alternative, equivalent, definitions of topological entropy, in the compact metric-space case, that are often easier to work with than covers. Bowen gave a definition also when X is not compact (see Bowen 1973; Chap. 7 in Walters 1981). (When X is not compact, the definitions need not coincide; e. g., Hasselblatt et al. (2005). And topologically equivalent metrics, but which are not uniformly equivalent, may give the same T different entropies (see p. 171 in Walters 1981).)

Metric Preliminaries

An ϵ -ball-cover comprises finitely many balls, all of radius ϵ . Since our space is compact, every cover $\mathcal{V}$ has a **Lebesgue number** $\epsilon > 0$. I. e., for each $z \in X$, the $\text{Bal}(z, \epsilon)$ lies entirely inside at least one $\mathcal{V}$ -patch. (In particular, there is an ϵ -ball-cover which *refines* $\mathcal{V}$.) Let $\text{LEB}(\mathcal{V})$ be the supremum of the Lebesgue numbers. Courtesy of Lemma 15f we can

Fix a “universal” list $\mathcal{V}^{(1)} \leq \mathcal{V}^{(2)} \leq \dots$, with $\mathcal{V}^{(k)}$ a $\frac{1}{k}$ ball – cover. For every $T : X \rightarrow X$, then, the $\lim_k h(T, \mathcal{V}^{(k)})$ computes $h_{\text{top}}(T)$.

An ϵ -Microscope Three notions are useful in examining a metric space (X, d) at scale ϵ . Subset $A \subset X$ is an ϵ -**separated-set**, if $d(z, z') \geq \epsilon$ for all distinct $z, z' \in A$. Subset $F \subset X$ is ϵ -**spanning** if $\forall x \in X, \exists z \in F$ with $d(x, z) < \epsilon$.

Lastly, a cover $\mathcal{V}$ is ϵ -small if $\text{Diam}(U) < \epsilon$, for each $U \in \mathcal{V}$.

You Take the High Road and I’ll Take the Low Road

There are several routes to computing top-ent, some via maximization, others, minimization. Our foregoing discussion computed $h_{\text{top}}(T)$ by a family of *sizes* $f_k(n) = f_k^T(n)$, depending on a

parameter k which specifies the fineness of scale. (In section “[Metric Preliminaries](#),” this k is an integer; in the original definition, an open cover.) Define two numbers:

$$\begin{aligned}\overline{L}^f(k) &:= \limsup_{n \rightarrow \infty} \frac{1}{n} \log f_k(n) \quad \text{and} \\ \underline{L}^f(k) &:= \liminf_{n \rightarrow \infty} \frac{1}{n} \log f_k(n).\end{aligned}\quad (22)$$

Finally, let $h^f(T) := \sup_k \overline{L}^f(k)$. If the limit exists in (22) then agree to write $L^f(k)$ for the common value. The A-K-M definition used the size $f_{\mathcal{V}}(n) := \text{Card}(\mathcal{V}_{[0..n]})$, where

$\text{Card}(\mathcal{W}) :=$ Minimum cardinality of
a subcover from $\mathcal{W}$.

Here are three metric-space sizes $f_{\in}(n)$:

$\text{Sep}(n, \varepsilon) :=$ Maximum cardinality of
a $d_n - \varepsilon$ - separated set.

$\text{Spn}(n, \varepsilon) :=$ Minimum cardinality of
a $d_n - \varepsilon$ - spanning set.

$\text{Cov}(n, \varepsilon) :=$ Minimum cardinality of
a $d_n - \varepsilon$ - small cover.

These use a list $(d_n)_{n=1}^{\infty}$ of progressively finer metrics on X , where

$$d_N(x, y) := \max_{j \in [0..N]} d(T^j(x), T^j(y)).$$

Theorem 16 (All-Roads-Lead-to-Rome Theorem) Fix ϵ and let $\mathcal{W}$ be any $d - \epsilon$ -small cover. Then

- (a) $\forall n: \text{Cov}(n, 2\epsilon) \leq \text{Spn}(n, \epsilon) \leq \text{Sep}(n, \epsilon) \leq \text{Card}(\mathcal{W}_{[0..n]})$.
- (b) Take a cover $\mathcal{V}$ and a $\delta < \text{LEB}(\mathcal{V})$. Then $\forall n: \text{Card}(\mathcal{V}_{[0..n]}) \leq \text{Cov}(n, \delta)$.
- (c) The limit $L^{\text{Cov}}(\epsilon) = \lim_n \frac{1}{n} \log \text{Cov}(n, \epsilon)$ exists in $[0.. \infty)$.
- (d) $h^{\text{Sep}}(T) = h^{\text{Spn}}(T) = h^{\text{Cov}}(T) = h^{\text{Card}}(T) \stackrel{\text{by defn}}{=} h_{\text{top}}(T)$.

Proof (a) Take $F \subset X$, a min-cardinality $d_n - \epsilon$ -spanning set. So $\cup_{z \in F} D_z = X$, where

$$D_z := d_n - \text{Bal}(z, \epsilon) \stackrel{\text{note}}{=} \bigcap_{j=0}^{n-1} T^{-j}(\text{Bal}(T^j z, \epsilon)).$$

This $\mathcal{D} := \{D_z\}_z$ is a cover, and it is $d_n - 2\epsilon$ -small. Thus, $\text{Cov}(n, 2\epsilon) \leq |\mathcal{D}| = |F|$. For any metric, a *maxima* ϵ -separated-set is automatically ϵ -spanning; adjoin a putative unspanned point to get a larger separated set.

Let A be a max-cardinality $d_n - \epsilon$ -separated set. Take C , a min-cardinality subcover of $\mathcal{W}_{[0..n]}$. For each $z \in A$, pick a C -patch $C_z \ni z$. Could some pair $x, y \in A$ pick the same C ? Well, write $C = \cap_{j=0}^{n-1} T^{-j}(W_j)$, with each $W_j \in \mathcal{W}$. For every $j \in [0..n)$, then,

$$d(T^j(x), T^j(y)) \leq \text{Diam}(W_j) < \epsilon.$$

Hence, $d_n(x, y) < \epsilon$; so $x = y$. Accordingly, the $z \mapsto C_z$ map is injective, whence $|A| \leq |C|$.

(b) Choose a min cardinality $d_n - \delta$ -small cover C . For each $C \in C$ and $j \in [0..n)$, the d -Diam($T^j C$) $< \delta$. So there is a $\mathcal{V}$ -patch $V_{C,j} \supset T^j(C)$. Hence

$$\mathcal{V}_{[0..n]} \stackrel{\text{note}}{\ni} \bigcap_{j=0}^{n-1} T^{-j}(V_{C,j}) \supset C.$$

Thus, $\mathcal{V}_{[0..n]} \preccurlyeq C$. So

$$\text{Card}(\mathcal{V}_{[0..n]}) \leq \text{Card}(C) \leq |C| = \text{Cov}(n, \delta).$$

(c) To upper-bound $\text{Cov}(k+l, \epsilon)$ let $\mathcal{V}$ and $\mathcal{W}$ be min-cardinality ϵ -small covers, respectively, for metrics d_k and d_l . Then $\mathcal{V} \cap T^l(\mathcal{W})$ is a ϵ -small for d_{k+l} . Consequently $\text{Cov}(k+l, \epsilon) \leq \text{Cov}(k, \epsilon) \cdot \text{Cov}(l, \epsilon)$. Thus, $n \mapsto \log(\text{Cov}(n, \epsilon))$ is subadditive.

(d) Pick a $\mathcal{V}$ from the list in section “[Metric Preliminaries](#),” choose some $2\epsilon < \text{LEB}(\mathcal{V})$ followed by an ϵ -small $\mathcal{W}$ from section “[Metric Preliminaries](#).” Pushing $n \rightarrow \infty$ gives

$$\begin{aligned}
 L^{\text{Card}}(\mathcal{V}) &\leq L^{\text{Cov}}(2\varepsilon) \\
 &\leq \bar{L}^{\text{Spn}}(\varepsilon) \leq \bar{L}^{\text{Sep}}(\varepsilon) \\
 &\leq \underline{L}^{\text{Spn}}(\varepsilon) \leq \underline{L}^{\text{Sep}}(\varepsilon) \\
 &\leq L^{\text{Card}}(\mathcal{W}). \quad (23)
 \end{aligned}$$

Now send $\mathcal{V}$ and $\mathcal{W}$ along the list in section “Metric Preliminaries.”

Pretension

Topological entropy takes its values in $[0, \infty]$. A useful corollary of (23) can be stated in terms of any Distance $(\cdot, \cdot)$ which topologizes $[0, \infty]$ as a compact interval.

For each continuous $T: X \rightarrow X$ on a compact metric-space, the Distance $(\underline{L}^{\text{Sep}}(\varepsilon), \bar{L}^{\text{Sep}}(\varepsilon))$, goes to zero as $\varepsilon \searrow 0$. Consequently, we can pretend that the

$$L^{\text{Sep}}(\varepsilon) = \lim_{n \rightarrow \infty} \frac{1}{n} \log(\text{Sep}(n, \varepsilon)) \quad (24)$$

limit exists, in arguments that subsequently send $\varepsilon \searrow 0$. Ditto for $\bar{L}^{\text{Spn}}(\varepsilon)$.

This will be used during the proof of the **Variational Principle**. But first, here are two entropy computations which illustrate the efficacy in having several characterizations of topological entropy.

$$h_{\text{top}}(\text{Isometry}) = 0$$

Suppose $(T: X, d)$ is a distance-preserving map of a compact metric-space. Fixing ϵ , a set is $d_n - \epsilon$ -separated *IFF* it is $d - \epsilon$ -separated. Thus, $\text{Sep}(n, \epsilon)$ does *not* grow with n . So each $\bar{L}^{\text{Sep}}(\epsilon)$ is zero.

Topological Markov Shifts

Imagine ourselves back in the days when computer data is stored on large reels of fast-moving magnetic tape. One strategy to maximize the density of binary data stored is to *not* put timing-marks (which take up space) on the tape. This has the defect that when the tape-writer writes, say, 577 consecutive 1-bits, then the tape-reader may erroneously count 578 copies of 1. We side-step this flaw by first encoding our data so as to avoid the $11^{577}1$ word, then writing to tape.

Generalize this to a finite alphabet Q and a finite list $\mathcal{F}$ of disallowed Q -words. Extend each word to a common length $K + 1$; now $\mathcal{F} \subset Q^{K+1}$

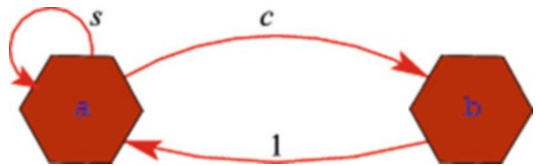
The resulting “ K -step **TMS (topological Markov shift)**” is the shift on the set of doubly- ∞ Q -names having no substring in $\mathcal{F}$. In the above magnetic-tape example, $K = 576$. Making it more realistic, suppose that some string of zeros, say $00^{574}0$, is also forbidden. (Perhaps the ϕ -bad-length, 574, is shorter than the 1-bad-length because, say, ϕ s take less tape-space than 1s and so – being written more densely – cause ambiguity sooner.) Extending to length 577, we get $2^3 = 8$ new disallowed words of form $00^{574}0b_1b_2b_3$.

We *recode* to a 1-step **TMS** (just called a **TMS** or a **subshift of finite type**) over the alphabet $P := Q^K$. Each outlawed Q -word $w_0w_1 \cdots w_K$ engenders a length-2 forbidden P word $(w_0, \dots, w_{K-1})(w_1, \dots, w_K)$. The resulting **TMS** is topologically conjugate to the original K -step. The *allowed* length-2 words can be viewed as the edges in a directed-graph and the set of points $x \in X$ is the set of doubly- ∞ paths through the graph. Once trivialities removed, this X is a Cantor set and the shift $T: X \rightarrow X$ is a homeomorphism (see Chaps. 2 and 6 in Lind and Marcus 1996).

The Golden Shift

As the simplest example, suppose our magnetic-tape is constrained by the Markov graph, Fig. 5 that we studied measure-theoretically in Fig. 4.

We want to store the text of *The Declaration of Independence* on our magnetic tape. Imagining that English is a stationary process, we’d like to encode English into this Golden **TMS** as efficiently as possible. We seek a shift-invariant



Entropy in Ergodic Theory, Fig. 5 Ignoring the labels on the edges, for the moment, the **Golden shift**, T , acts on the space of doubly-infinite paths through this graph. The space can be represented as a subset $X_{\text{Gold}} \subset \{a, b\}^{\mathbb{Z}}$, namely, the set of sequences with no two consecutive b letters

measure μ on X_{Gold} of *maximum entropy*, should such exist.

View $P = \{a, b\}$ as the time-zero partition on X_{Gold} ; that is, name $x = \dots x_{-1}x_0x_1x_2\dots$, is in atom b *IFF* letter x_0 is “ b .” Any measure μ gives conditional probabilities

$$\begin{aligned} \mu(a|a) &=: s, & \mu(b|a) &=: c, \\ \mu(a|b) &\stackrel{\text{note}}{=} 1, & \mu(b|b) &\stackrel{\text{note}}{=} 0. \end{aligned}$$

But recall, $h(T) = \mathcal{H}(P_1 | P_{[-\infty \dots 0)}) \leq \mathcal{H}(P_1 | P_0)$. So among all measures that make the conditional distribution $P | a$ equal (s, c) , the *unique* one maximizing entropy is the (s, c) -Markov process. Its entropy, derived in (14), is

$$\begin{aligned} f(s) &:= \frac{1}{2-s} \mathcal{H}(s, 1-s) \\ &= \frac{-1}{2-s} [s \log(s) + (1-s) \log(1-s)]. \end{aligned} \quad (25)$$

Certainly $f(0) = f(1) = 0$, so f 's maximum occurs at the (it turns out) *unique* point $\hat{s}$ where the derivative $f'(\hat{s})$ equals zero. This $\hat{s} = (-1 + \sqrt{5})/2$. Plugging in, the maximum entropy supportable by the Golden Shift is

$$\begin{aligned} \text{MaxEnt} &= \frac{2}{5 - \sqrt{5}} \left[\frac{-1 + \sqrt{5}}{2} \log \left(\frac{2}{-1 + \sqrt{5}} \right) \right. \\ &\quad \left. + \frac{3 - \sqrt{5}}{2} \log \left(\frac{2}{3 - \sqrt{5}} \right) \right]. \end{aligned} \quad (26)$$

Exponentiating, the number of μ -typical n -names grows like G^n , where

$$G := \left[\frac{2}{-1 + \sqrt{5}} \right]^{\frac{-1 + \sqrt{5}}{5 - \sqrt{5}}} \left[\frac{2}{3 - \sqrt{5}} \right]^{\frac{3 - \sqrt{5}}{5 - \sqrt{5}}}. \quad (27)$$

This expression looks unpleasant to simplify – it is not even obviously an algebraic number – and yet topological entropy will reveal its familiar nature. This is because the Variational Principle (proved in the next section) says that the top-ent of a system is the supremum of measure-entropies supportable by the system³.

Top-Ent of the Golden Shift

For a moment, let's work more generally on an arbitrary subshift (a closed, shift-invariant subset) $X \subset Q^{\mathbb{Z}}$, where Q is a finite alphabet. Here, the transformation is always the shift – but the *space* is varying – so agree to refer to the top-ent as $h_{\text{top}}(X)$. Let $\text{Names}_X(n)$ be the number of distinct words in the set $\{x|_{[0..n]} \mid x \in X\}$. Note that a metric inducing the product-topology on $Q^{\mathbb{Z}}$ is

$$d(x, x') := \frac{1}{1 + |m|}, \quad (28)$$

for the smallest $|m|$ with $x_m \neq x'_m$.

Lemma 17 *Consider a subshift X . Then the*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log(\text{Names}_X(n))$$

exists in $[0, \infty]$, and equals $h_{\text{top}}(X)$.

Proof With $\varepsilon \in (0, 1)$ fixed, two n -names are $d_n - \varepsilon$ -separated *IFF* they are not the same name. Hence, $\text{Sep}(n, \varepsilon) = \text{Names}_X(n)$.

To compute $h_{\text{top}}(X_{\text{Gold}})$, declare that a word is “golden” if it appears in some $x \in X_{\text{Gold}}$. Each $[n + 1]$ -golden word ending in a has form wa , where w is n -golden. An $[n + 1]$ -golden word ending in b , must end in ab and so has form wab , where w is $[n - 1]$ -golden. Summing up, $\text{Names}_{X_{\text{Gold}}}(n + 1) = \text{Names}_{X_{\text{Gold}}}(n) + \text{Names}_{X_{\text{Gold}}}(n - 1)$. This is the Fibonacci recurrence, and indeed, these are *the* Fibonacci numbers, since $\text{Names}_{X_{\text{Gold}}}(0) = 1$ and $\text{Names}_{X_{\text{Gold}}}(1) = 2$. Consequently, we have that

$$\text{Names}_{X_{\text{Gold}}}(n) \sim \text{Const} \cdot \lambda^n,$$

where $\lambda = \frac{1 + \sqrt{5}}{2}$ is the Golden Ratio. So the sesquipedalian number G from (27) is simply λ , and $h_{\text{top}}(X_{\text{Gold}}) = \log(\lambda)$. Since $\log(\lambda) \approx 0.694$, each thousand bits written on tape (subject to the “no

³A popular computer-algebra-system was not, at least under our in-expert tutelage, able to simplify this. However, once top-ent gave the correct answer, the software was able to detect the *equality*.

bb substrings” constraint) can carry at most 694 bits of information.

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Top-Ent of a General TMS

A (finite) digraph G engenders a **TMS** $T: X_G \rightarrow X_G$, as well as a $\{0, 1\}$ -valued adjacency matrix $\mathbf{A} = \mathbf{A}_G$, where $a_{i,j}$ is the number of directed-edges from state i to j . (Here, each $a_{i,j}$ is 0 or 1.) The (i, j) -entry in power $\mathbf{A}^n$ is automatically the number of length- n paths from i to j . Employing the matrix-norm $\|\mathbf{M}\| := \sum_{i,j} |m_{i,j}|$, then,

$$\|\mathbf{A}\|^n = \text{Names}_X(n).$$

Happily Gelfand’s formula (see 10.13 in Rudin 1973 or Spectral_radius in Wikipedia) applies: For an arbitrary (square) complex matrix,

$$\lim_{n \rightarrow \infty} \|\mathbf{A}^n\|^{\frac{1}{n}} = \text{SpecRad}(\mathbf{A}). \quad (29)$$

This right hand side, the *spectral radius* of $\mathbf{A}$, means the maximum of the absolute values of $\mathbf{A}$ ’s eigenvalues. So the top-ent of a **TMS** is thus the

$$\begin{aligned} h_{\text{top}}(X_G) &= \text{SpecRad}(\mathbf{A}_G) \\ &:= \text{Max}\{|e| \mid e \text{ is an eigenvalue of } \mathbf{A}_G\}. \end{aligned} \quad (30)$$

The (a, b)-adjacency matrix of Fig. 5 is

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix},$$

whose eigenvalues are λ and $\frac{1}{\lambda}$.

Labeling Edges

Interpret $(s, c, 1)$ simply as edge-labels in Fig. 5. The set of doubly - ∞ paths can also be viewed as a subset $Y_{\text{Gold}} \subset \{s, c, 1\}^{\mathbb{Z}}$, and it too is a **TMS**. The shift on Y_{Gold} is conjugate (topologically isomorphic) to the shift on X_{Gold} , so they *a fortiori* have the same top-ent, $\log(\lambda)$. The $(s, c, 1)$ -adjacency matrix is

Its $|\cdot|$ -largest eigenvalue is still λ , as it must.

Now we make a new graph. We modify Fig. 5 by manufacturing a total of two s -edges, seven c -edges, and three edges $1_1, 1_2, 1_3$. Give these $2 + 7 + 3$ edges twelve *distinct* labels. We *could* compute the resulting **TMS**-entropy from the corresponding 12×12 adjacency matrix. Alternatively, look at the (a, b)-adjacency matrix

$$\mathbf{A} := \begin{bmatrix} 2 & 7 \\ 3 & 0 \end{bmatrix}.$$

The roots of its characteristic polynomial are $1 \pm \sqrt{22}$. Hence, h_{top} of this 12-symbol **TMS** is $\log(1 + \sqrt{22})$.

The Variational Principle

Let $\mathcal{M} := \mathcal{M}(X, d)$ be the set of Borel probability measures, and $\mathcal{M}(T) := \mathcal{M}(X, d, T)$ the set of T -invariant $\mu \in \mathcal{M}$. Assign

$$\text{EntSup}(T) := \sup\{h_{\mu}(T) \mid \mu \in \mathcal{M}(T)\}$$

Theorem 18 (Variational Principle (Goodson)) $\text{EntSup}(T) = h_{\text{top}}(T)$.

This says that top-ent is the top entropy – *if* there is a measure μ which realizes the supremum. There doesn’t have to be. Choose a sequence of metric-systems (Y_k, m_k, S_k) whose entropies *strictly* increase $h_{\text{top}}(S_k) \nearrow L$ to some limit in $(0, \infty]$. Let $(Y_{\infty}, m_{\infty}, S_{\infty})$ be the identity-map on a 1-point space. Define a new system (X, d, T) , where $X := \bigsqcup_{k \in [1 \dots \infty]} Y_k$. Have $T(x) := S_k(x)$, for the unique k with $Y_k \ni x$. As for the metric, on Y_k let d be a scaled version of m_k , so that the d -Diam (Y_k) is less than $1/2^k$. Finally, for points in *distinct* components, $x \in Y_k$ and $z \in Y_{\ell}$, decree that $d(x, z) := |2^{-k} - 2^{-\ell}|$. Our T is continuous, and is a homeomorphism if each of the S_k is. Certainly $h_{\text{top}}(T) = L > h_{\text{top}}(S_k)$, for every $k \in [1 \dots \infty]$.

If L is *finite* then there is *no* measure μ of maximal entropy; for μ must give mass to *some* Y_k ; this pulls the entropy below L , since there are no compensatory components with entropy exceeding L .

In contrast, when $L = \infty$ then there *is* a maximal-entropy measure (put mass $1/2^j$ on some component Y_{k_j} , where $k_j \nearrow \infty$ swiftly); indeed, there are continuum-many maximal-entropy measures. *But* there is no *ergodic* measure of maximal entropy. (The ergodic measures are the extreme points of $\mathcal{M}(T)$; call them $\mathcal{M}_{\text{Erg}}(T)$. This $\mathcal{M}(T)$ is the set of barycenters obtained from Borel probability measures on $\mathcal{M}_{\text{Erg}}(T)$ (see Krein-Milman theorem, Choquet_theory in Wikipedia). In this instance, what explains the failure to have an *ergodic* maximal-entropy measure? Let μ_k be an invariant ergodic measure on Y_k . These measures *do* converge to the one-point (ergodic) probability measure μ_∞ on Y_∞ . But the map $\mu \mapsto h_\mu(T)$ is not continuous at μ_∞ .)

For a concrete $L = \infty$ example, let S_k be the shift on $[1 \dots k]^\mathbb{Z}$ (see Buzzi and Ruelle 2006; Denker 1976; Misiurewicz 1973 for measures of maximal entropy.)

Topology on $\mathcal{M}$

Let's arrange our tools for establishing the Variational Principle. The argument will follow Misiurewicz's proof, adapted from the presentations in Brin and Stuck (2002) and Walters (1981).

Equip $\mathcal{M}$ with the **weak-** * topology. (This metrizable topology makes $\mathcal{M}$ compact. Always, $\mathcal{M}(T)$ is a nonvoid compact subset (see "Measure-Preserving Systems").) That is, measure $\alpha_L \rightarrow \mu$ *IFF* $\int f d\alpha_L \rightarrow \int f d\mu$, for each continuous $f : X \rightarrow \mathbb{R}$. An $A \subset X$ is μ -**nice** if its topological boundary $\partial(A)$ is μ -null. And a *partition* is μ -nice if each atom is.

Proposition 19 *If $\alpha_L \rightarrow \mu$ and $A \subset X$ is μ nice, then $\alpha_L(A) \rightarrow \mu(A)$.*

Proof Define operator $\mathcal{U}(D) := \limsup_L \alpha_L(D)$. It suffices to show that $\mathcal{U}(A) \leq \mu(A)$. For since A^c is

μ -nice too, then $\mathcal{U}(A^c) \leq \mu(A^c)$. Thus, $\lim_L \alpha_L(A)$ exists and equals $\mu(A)$. Because $C := \overline{A}$ is closed, the continuous functions $f_N \searrow \mathbf{1}_C$ pointwise, where

$$f_N(x) := 1 - \text{Min}(N \cdot d(x, C), 1).$$

By the Monotone Convergence theorem, then,

$$\int f_N d\mu \xrightarrow{N} \mu(C).$$

And $\mu(C) = \mu(A)$, since A is *nice*. Fixing N , then, it suffices to establish $\mathcal{U}(A) \leq \int f_N d\mu$. But f_N is continuous, so

$$\begin{aligned} \int f_N d\mu &= \limsup_{L \rightarrow \infty} \int f_N d\alpha_L \\ &\geq \limsup_{L \rightarrow \infty} \int \mathbf{1}_A d\alpha_L = \mathcal{U}(A). \end{aligned}$$

Corollary 20 *Suppose $\alpha_L \rightarrow \mu$, and partition P is μ -nice. Then $\mathcal{H}_{\alpha_L}(P) \rightarrow \mathcal{H}_\mu(P)$.*

The **diameter** of partition P is $\text{Max}_{A \in P} \text{Diam}(A)$.

Proposition 21 *Take $\mu \in \mathcal{M}$ and $\varepsilon > 0$. Then there exists a μ -nice partition with $\text{Diam}(P) < \varepsilon$.*

Proof Centered at an x , the uncountably many balls $\{\text{Bal}(x, r) \mid r \in (0, \varepsilon)\}$ have disjoint boundaries. So all but countably many are μ -nice; pick one and call it B_x . Compactness gives a finite nice cover, say, $\{B_1, \dots, B_7\}$, at different centers. Then the partition $P := (A_1, \dots, A_7)$ is nice, where $A_k := B_k \setminus \bigcup_{j=1}^{k-1} B_j$. (For any two sets $B, B' \subset X$, the union $\partial B \cup \partial B'$ is a superset of the three boundaries $\partial(B \cup B')$, $\partial(B \cap B')$, $\partial(B \setminus B')$.)

Here is a consequence of Jensen's inequality.

Lemma 22 (Distropy-Averaging Lemma) *For $\mu, \nu \in \mathcal{M}$, a partition R , and a number $t \in [0, 1]$,*

$$t\mathcal{H}_\mu(R) + (1-t)\mathcal{H}_\nu(R) \leq \mathcal{H}_{t\mu+(1-t)\nu}(R).$$

Proof of the Variational Principle As usual we divide the proof into 2 parts. $\text{EntSup}(T) \geq h_{\text{top}}(T)$ and $\text{EntSup}(T) \leq h_{\text{top}}(T)$.

Strategy for $\text{EntSup}(T) \geq h_{\text{top}}(T)$. Choose an $\varepsilon > 0$. For $L = 1, 2, 3, \dots$, take a maximal (L, ε) -separated-set $F_L \subset X$, then define

$$\mathbf{F} = \mathbf{F}_\varepsilon := \limsup_{L \rightarrow \infty} \frac{1}{L} \log(|F_L|).$$

Let $\varphi_L()$ be the equi-probable measure on F_L ; each point has weight $1/|F_L|$. The desired invariant measure μ will come from the Cesàro averages

$$\alpha_L := \frac{1}{L} \sum_{\ell \in [0..L)} T^\ell \varphi_L,$$

which get more and more invariant.

Lemma 23 *Let μ be any weak-* accumulation point of the above $\{\alpha_L\}_1^\infty$. (Automatically, μ is T -invariant.) Then $h_\mu(T) \geq \mathbf{F}$. Indeed, if Q is any μ -nice partition with $\text{Diam}(Q) < \varepsilon$, then $h_\mu(T, Q) \geq \mathbf{F}$.*

Tactics As usual, $Q_{[0..N)}$ means $Q_0 \vee Q_1 \vee \dots \vee Q_{N-1}$. Our goal is to show:

$$\forall N : \quad \mathbf{F} \stackrel{?}{\leq} \frac{1}{N} \mathcal{H}_\mu(Q_{[0..N)}). \quad (31)$$

Fix N and $P := Q_{[0..N)}$, and a $\delta > 0$. It suffices to verify: $\forall_{\text{large } L} L \gg N$,

$$\frac{1}{L} \log(|F_L|) \stackrel{?}{\leq} \delta + \frac{1}{N} \mathcal{H}_{\alpha_L}(P), \quad (32)$$

since this and Corollary 20 will prove (31): Pushing $L \rightarrow \infty$ along the sequence that produced μ essentially sends LHS(32) to $\mathbf{F}$, courtesy (24). And RHS(32) goes to $\delta + \frac{1}{N} \mathcal{H}_\mu(P)$, by Corollary 20, since P is μ -nice. Descending $\delta \searrow 0$, hands us the needed (31).

Remark 24 The idea in the following proof is to mostly fill interval $[0 \dots L)$ with N -blocks, starting with a offset $K \in [0 \dots N)$. Averaging over the offset will create a Cesàro average over each N -block. Averaging over the N -blocks will allow us to compute distropy with respect to the averaged measure, α_L .

Proof (of (32)) Since L is fixed, agree to use φ for the φ_L probability measure. Our d_L - ε -separated set F_L has at most *one* point in any given atom of $Q_{[0..L)}$, thereupon

$$\log(|F_L|) = \mathcal{H}_\varphi(Q_{[0..L)}).$$

Regardless of the “offset” $K \in [0 \dots N)$, we can always fit $C := \lfloor \frac{L-N}{N} \rfloor$ many N -blocks into $[0..L)$. Denote by $\mathcal{G}(K) := [K \dots K + CN)$, this union of N -blocks, the **good** set of indices. Let $\mathcal{B}(K) := [0..L) \setminus \mathcal{G}(K)$ be the **bad** index-set. Therefore,

$$\mathcal{H}_\varphi(Q_{[0..L)}) \leq \overbrace{\mathcal{H}_\varphi\left(\bigvee_{j \in \mathcal{B}(K)} Q_j\right)}^{\text{Bad}(K)} + \underbrace{\mathcal{H}_\varphi\left(\bigvee_{i \in \mathcal{G}(K)} Q_j\right)}_{\text{Good}(K)}. \quad (33)$$

Certainly $\text{Bad}(K) \leq 3N \log(|Q|)$. So

$$\frac{1}{NL} \sum_{K \in [0..N)} \text{Bad}(K) \leq \frac{3N}{L} \log(|Q|).$$

This is less than δ , since L is large. Applying $\frac{1}{NL} \sum_{K \in [0..N)}$ to (28) now produces

$$\frac{1}{L} \log(|F_L|) \leq \delta + \frac{1}{NL} \sum_K \text{Good}(K). \quad (34)$$

Note

$$\bigvee_{j \in \mathcal{G}(K)} T^j(Q) = \bigvee_{c \in [0 \dots C) T^{K+cN}(P)}.$$

So

$$\text{Good}(K) \leq \sum_c \mathcal{H}_\varphi(T^{K+cN}P).$$

This latter, by definition, equals $\sum_c \mathcal{H}_{T^{K+cN}(\varphi)}(P)$. We conclude that

$$\begin{aligned} \frac{1}{NL} \sum_K \text{Good}(K) &\leq \frac{1}{NL} \sum_K \sum_c \mathcal{H}_{T^{K+cN}(\varphi)}(P) \\ &\leq \frac{1}{NL} \sum_{\ell \in [0 \dots L)} \mathcal{H}_{T^\ell \varphi}(P), \\ &\quad \text{by adjoining a few translates of } P, \\ &\leq \frac{1}{N} \mathcal{H}_{\alpha_L}(P), \\ &\quad \text{by the Distropy – Averaging Lemma 22,} \end{aligned}$$

since α_L is the average $\frac{1}{L} \sum_\ell T^\ell \varphi$. Thus, (34) implies (32), our goal.

Proof of $\text{EntSup}(T) \leq h_{\text{top}}(T)$. Fix a T -invariant measure μ . For partition $Q = (B_1, \dots, B_K)$, choose a compact set $A_k \subset B_k$ with $\mu(B_k \setminus A_k)$ small. (This can be done, since μ is automatically a regular measure (Rudin 1973).) Letting $D := [\bigcup_i A_i]^c$ and $P := (D, A_1, \dots, A_K)$, we can have made $\mathcal{H}(P|Q)$ as small as desired. Courtesy of Lemma 8b, then, we only need consider partitions of the form that P has.

Open-cover $\mathcal{V} = (U_1, \dots, U_K)$ has patches $U_k := D \cup A_k$. What atoms of, say, $P_{[0,1,2]}$, can the intersection $U_9 \cap T^{-1}(U_2) \cap T^{-2}(U_5)$ touch? Only the eight atoms

$$(D \text{ or } A_9) \cap T^{-1}(D \text{ or } A_2) \cap T^{-2}(D \text{ or } A_5).$$

Thus, $\#P_{[0 \dots n]} \leq 2^{n \cdot \# \mathcal{V}_{[0 \dots n]}}$. Here $\#()$ counts the number of nonvoid atoms/patches.) So

$$\begin{aligned} \frac{1}{n} \mathcal{H}_\mu(P_{[0 \dots n]}) &\leq 1 + \frac{1}{n} \log(\# \mathcal{V}_{[0 \dots n]}) \\ &\leq 1 + 1 + h_{\text{top}}(T); \end{aligned}$$

this last inequality, when n is large. The upshot: $h_\mu(T) \leq 2 + h_{\text{top}}(T)$.

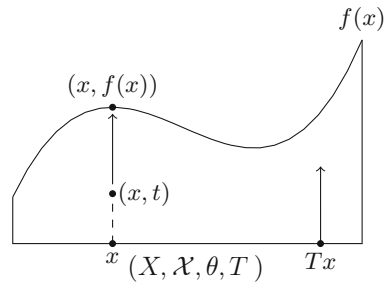
Applied to a power T^ℓ , this asserts that $h_\mu(T^\ell) \leq 2 + h_{\text{top}}(T^\ell)$. Thus, $h_\mu(T) \leq \frac{2}{\ell} + h_{\text{top}}(T)$, using Lemma 9d and using Lemma 15g. Now let $\ell \rightarrow \infty$.

Entropy of a Flow($\mathbb{R}$ -Action)

Historically study of dynamical systems started with the continuous $\mathbb{R}$ -action governed by Newton's law. We denote the system by $(\Omega, \mathcal{G}, v, T_t)$ where $T_{t_2}(T_{t_1}(x)) = T_{t_2+t_1}(x)$. The geodesic flows and horocycle flows in homogeneous dynamics are well-known example of $\mathbb{R}$ -actions. For each given t_0 , $(\Omega, \mathcal{G}, v, T_{t_0})$ gives rise to a $\mathbb{Z}$ -action. We define $h(T_t) = h(T_1)$, entropy of time 1 map. It is not hard to see that $h(T_{ct}) = ch(T_t)$ for every $c \in \mathbb{R}$. In particular, If c is an integer, then $h(T_c) = h(T_1) = ch(T_t)$ (see Lemma 9d).

Ambrose (1941) has shown that every ergodic flow $(\Omega, \mathcal{G}, v, T_t)$ can be represented as a flow with a transformation on X and a ceiling function $f(x)$, which is measurable with respect to $\mathcal{X}$. When we denote the base by $(X, \mathcal{X}, \mu, T)$, then the flow T_t acts as follows: each point $\omega = (x, t)$ moves straight up at unit speed until it hits $(x, f(x))$ which is identified as $(Tx, 0)$. Then it continues up at unit speed. The base $(X, \mathcal{X}, \mu, T)$ is called a cross section of the flow. It is clear that the flow has positive entropy IFF the base $(X, \mathcal{X}, \mu, T)$ has positive entropy. Unpredictability of the base is inherited by the flow.

The flow has the following structure (Fig. 6).



Entropy in Ergodic Theory, Fig. 6 In the representation of $(\Omega, \mathcal{G}, v, T_{t_0})$, $(x, f(x))$ is identified as $(Tx, 0)$

- (i) $\Omega = \{(x, t) : x \in X, 0 \leq t \leq f(x)\}$.
- (ii) The σ -algebra $\mathcal{G}$ is the product σ -algebra of X and Lebesgue on the time axis. Also the measure $\nu = \mu \times \text{Lebesgue measure}$.
- (iii) $\int f(x) d\mu = 1$.

Rudolph strengthened the representation showing that the ceiling function $f(x)$ can have only two irrationally related values, say τ_0 and τ_1 where $\frac{\tau_0}{\tau_1}$ is irrational. Let $Q = \{Q_0, Q_1\}$ be the partition of X where $Q_0 = \{x : f(x) = \tau_0\}$. Let (X, P, μ, T) be the base, where $P = \{P_0, P_1, \dots, P_{k-1}\}$ is a generating partition. We denote $R = P \vee Q$. The partition R has at most $2k$ atoms. Let $R = \{R_1, \dots, R_l\}$. Let $\hat{R} = \{\hat{R}_1, \dots, \hat{R}_l\}$ be the partition of Ω , where $\hat{R}_i = \{(x, t) : x \in R_i, 0 \leq t < f(x)\}$.

Theorem 25 *The partition $\hat{R}$ is a generating partition of Ω under any ergodic transformation T_γ .*

Proof For the simplicity of argument, we assume that $\gamma < \min\{\tau_0, \tau_1\}$. Let (x_1, t_1) and (x_2, t_2) be two distinct points in Ω . If x_1 and x_2 belong to different atoms of R , then so do (x_1, t_1) and (x_2, t_2) of $\hat{R}$.

Suppose (x_1, t_1) and (x_2, t_2) is in $\hat{R}_i$.

- (i) If $x_1 \neq x_2$ and $t_1 = t_2$, then different R -names of x_1 and x_2 give rise to different $\hat{R}$ -names of (x_1, t_1) and (x_2, t_2) .
- (ii) If $x_1 = x_2$, then (x_1, t_1) and (x_2, t_2) are in the same orbit of T_γ . We assume $t_1 < t_2$ and let $\delta = \frac{t_2 - t_1}{2}$. Let

$$A = \{(x, t) : x \in R_{i_0}, T^{-1}x \in R_{i_1} (i_0 \neq i_1), 0 \leq t \leq \delta\}.$$

Since $\nu A > 0$ and T_γ is ergodic, there exists n such that $T_\gamma^n(x, t_2) \in A$. Hence,

$$\hat{R}(T_\gamma^n(x, t_2)) \in \hat{R}_{i_0} \text{ and } \hat{R}(T_\gamma^n(x, t_1)) \in \hat{R}_{i_1}.$$

- (iii) When $x_1 \neq x_2$ and $t_1 \neq t_2$, a little more thought leads to the above case either (i) or (ii).

Consider the following flow of Totoki type (Shields 1973).

Let (X, P, μ, T) be an independent process and

$$f(x) = \begin{cases} \tau_0 & \text{if } x \in P_0 \\ \tau_1 & \text{if } x \in P_1 \end{cases}$$

where $\{\tau_0, \tau_1\}$ are irrationally related. It is known that the flow is a Bernoulli flow. That is, for every t_0 , T_{t_0} is Bernoulli. Hence, by the isomorphism theorem of Bernoulli transformation (Entropy determines a Bernoulli transformation up to isomorphism), we can realize every Bernoulli transformation via a Bernoulli flow.

Entropy of $\mathbb{Z}^d$ -Actions

By a $\mathbb{Z}^d$ -action, σ , we mean commuting transformation $\sigma = \{T_1, T_2, \dots, T_d\}$. An element $\vec{n} = (n_1, n_2, \dots, n_d) \in \mathbb{Z}^d$ corresponds to $T_1^{n_1} T_2^{n_2} \dots T_d^{n_d} = \sigma^{\vec{n}}$ acting on $(X, \mathcal{X}, \mu)$. Note that study of $\mathbb{Z}^d$ -actions is a stepping stone of amenable group-actions. For simplicity of notations we assume $d = 2$. (The following definitions can be easily generalized to $\mathbb{Z}^d$ -actions, $d \geq 2$.)

Like a transformation, we define

$$h(\sigma, P) = \lim_{n \rightarrow \infty} \frac{1}{n^2} \mathcal{H} \left(\bigvee_{(i,j) \in R_n} \sigma^{-(i,j)} P \right)$$

where $R_n = [0, n] \times [0, n]$ is a square and

$$h(\sigma) = \sup_P h(\sigma, P).$$

It is not hard to see that

$$h(\sigma) = h(\sigma, \tilde{P})$$

where $\tilde{P}$ is a generating partition under σ . That is, $\tilde{P}$ satisfies

$$\lim_{n \rightarrow \infty} \bigvee_{-n \leq i, j \leq n} \sigma^{-(i,j)} \tilde{P} = \mathcal{X}.$$

We also have that

$$h(\sigma, P) = \mathcal{H}(P|P^-)$$

where $P^- = V_{(i,j)} < (0, 0)\sigma^{-(i,j)}P$ and “ $<$ ” is a lexicographic order.

In more general amenable groups, the sequence of square R_n is replaced by a sequence of Følner sets. Existence of Følner sequence guarantees the existence of invariant probability measure under the action (Ornstein and Weiss 1987).

If $h(\sigma, P) > 0$, then $\mathcal{H}(V_{(i,j) \in R_n} \sigma^{-(i,j)}P)$ is in the order of n^2 . We have the following theorem.

Theorem 26 *If $h(\sigma) > 0$, then $h(T_1^i T_2^j) = \infty \forall (i, j) \in \mathbb{Z}^2 \setminus \{(0, 0)\}$.*

Proof For simplicity we will prove for the case $(i, j) = (1, 0)$. Let P be a generating partition under σ . We note that a generating partition P under σ is not necessarily a generating partition for $T_1 = T^{(1,0)}$. Let P_{-k}^m denote the partition $V_{j=-k}^m T_2^j P$ and let $X_m = V_{-\infty}^\infty T_1^i (P_{-m}^m)$. Note that $\lim_{m \rightarrow \infty} X_m \nearrow X$.

$$\begin{aligned} h(T_1) &= h(T, P_{-m}^m) \\ &= \lim_{m \rightarrow \infty} \mathcal{H}\left(P_{-m}^m \mid \bigvee_{i=-\infty}^{-1} T_1^{-i} P_{-m}^m\right) \\ &= \lim_{m \rightarrow \infty} \sum_{k=-m}^m \mathcal{H}\left(T_2^k P \mid \left(\bigvee_{i=-\infty}^{-1} T_1^{-i} P_{-m}^m\right) \vee \left(\bigvee_{j=-m}^{k-1} T_2^{-j} P\right)\right) \\ &= \lim_{m \rightarrow \infty} \sum_{k=-m}^m \mathcal{H}\left(P \mid T_2^{-k} \left(\left(\bigvee_{i=-\infty}^{-1} T_1^{-i} P_{-m}^m\right) \vee \left(\bigvee_{j=-m}^{k-1} T_2^{-j} P\right)\right)\right) \end{aligned}$$

Since $P^- = V_{(i,j)} < (0, 0)\sigma^{-(i,j)}P$ is finer algebra than $T_2^{-k} \left(\left(\bigvee_{i=-\infty}^{-1} T_1^{-i} P_{-m}^m \right) \vee \left(\bigvee_{j=-m}^{k-1} T_2^{-j} P \right) \right)$, by Conditional-Distropy Fact (c), we have

$$\begin{aligned} \mathcal{H}\left(P \mid T_2^{-k} \left(\left(\bigvee_{i=-\infty}^{-1} T_1^{-i} P_{-m}^m \right) \vee \left(\bigvee_{j=-m}^{k-1} T_2^{-j} P \right) \right)\right) \\ > h(\sigma, P) \end{aligned}$$

for each m . Hence

$$\begin{aligned} h(T_1) &\geq \mathcal{H}(P_{-m}^m \mid \bigvee_{i=-\infty}^{-1} T_1^{-i} P_{-m}^m) \\ &\geq \sum_{k=-m}^m h(\sigma, P) \\ &= (2m+1)h(\sigma, P) \text{ for all } m \end{aligned}$$

Hence, $h(T_1) = \infty$. Similar reasoning for topological entropy gives that if $h_{\text{top}}(\sigma) > 0$, then $h_{\text{top}}(\sigma^{(i,j)}) = \infty$ for all $(i, j) \in \mathbb{Z}^2$.

Complexity of Zero Entropy Actions

As the study of dynamical systems expands to bigger group actions, entropy zero systems arise more naturally with diverse growth rates of orbits. We consider the following $\mathbb{Z}^2$ -action $\sigma = \{T_1, T_2\}$ where $h(\sigma) = 0$.

1. Let $\sigma = \{T_1, T_2\}$ be an action where T_1 and T_2 are irrational rotations by α_1 and α_2 respectively on a circle. Then $h(T_1) = h(T_2) = 0$. Hence, $h(\sigma) = 0$.
2. We have the following $\mathbb{Z}^2$ -subshift called three dot model.

Let $X = \{(x_{(i,j)} : x_{i+1,j} + x_{i,j} + x_{i,j+1} \equiv 0 \pmod{2})\} \subset \{0, 1\}^{\mathbb{Z}^2}$. It is not hard to see that

$$(a) \lim_{n \rightarrow \infty} \frac{1}{2n} \mathcal{H}_{\text{top}}(V_{(i,j) \in R_n} \sigma^{-(i,j)}P) = 1.$$

Hence, $h_{\text{top}}(\sigma) = 0$.

(b) For each P_{-m}^m ,

$$\begin{aligned} \mathcal{H}_{\text{top}}\left(P_{-m}^m \mid \bigvee_{i=-\infty}^{-1} T_1^{-i} P_{-m}^m\right) \\ = \mathcal{H}_{\text{top}}\left(T_2^m P \mid \bigvee_{i=-\infty}^{-1} T_1^{-i} P_{-m}^m \vee P_{-m}^{m-1}\right) = 1. \end{aligned}$$

Hence, $h_{\text{top}}(T_1) = 1$. Likewise $h_{\text{top}}(T_2) = 1$.

There exists an obvious invariant measure μ on X such that each atom of $V_{(i,j) \in [0, m) \times [0, n)} \sigma^{-(i,j)}P$ has measure $2^{-(m+n)}$. With respect to μ , we have $h(T_1) = h(T_2) = 1$. Hence, $h(\sigma) = 0$.

More generally, if Y is a shift space with the left shift T and ϕ is a continuous map from Y to itself, then it is well known that ϕ is a block code since ϕ is uniformly continuous. If ϕ is not invertible, then we use the “natural extension method” to extend the space Y to $X \in \{0, 1\}^{\mathbb{Z}}$ as in the above example (2). If Y has an invariant measure and ϕ is measure-preserving, then X carries an invariant measure.

To investigate the complexity of entropy zero $\mathbb{Z}^2$ -actions, J. Milnor (1988) introduced the notion of directional entropy, $h(\vec{v})$ for $\vec{v} = (x, y) \in \mathbb{R}^2$ as follows. Let P be a partition of X . We denote the partition $\sigma^{-(i,j)}P$ by $P^{-(i,j)}$

$$h(\vec{v}, P) = \overline{\lim}_{m \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} \frac{1}{t} \mathcal{H} \left(\bigvee_{(i,j) \in \mathcal{P}(m, \vec{v}, t)} P^{-(i,j)} \right)$$

where $\mathcal{P}(m, \vec{v}, t) = \{(i, j): 0 \leq j \leq [ty], -m - j \leq i \leq m + j\}$ and if $y = 0$, then $\mathcal{P}(m, \vec{v}, t) = \{(i, j): -m \leq j \leq m, 0 \leq i \leq [tx]\}$. That is, $\mathcal{P}(m, \vec{v}, t)$ is a parallelogram with width $[m]$ along the vector $\vec{v}$. Both limsups in the definition of $h(\vec{v}, P)$ are equal to limits. We define $h(\vec{v}) = \sup_P h(\vec{v}, P)$. If P is a generating partition under σ , then $h(\vec{v}) = h(\vec{v}, P)$ since $\mathcal{X}_m = \bigvee_{t=-\infty}^{\infty} \bigvee_{(i,j) \in \mathcal{P}(m, \vec{v}, t)} P^{-(i,j)} \nearrow \mathcal{X}$.

Directional entropy has the following properties.

- (i) $h(c\vec{v}) = ch(\vec{v})$.
- (ii) $h(T_1^i T_2^j \vec{v}) = h(\vec{v})$ where $\vec{v} = (i, j) \in \mathbb{Z}^2$
- (iii) $h(\vec{v})$ is not continuous in general for $\|\vec{v}\| = 1$.

We note that

- (a) Example (1) satisfies that $h(\vec{v}) = 0 \forall \vec{v} \in S^1$.
- (b) Example (2) satisfies that $h(\vec{v}) > 0 \forall \vec{v} \in S^1$.

However, an example with the following properties is known.

- (i) $h(\sigma^{(1,0)}) > 0$.
- (ii) $h(\sigma^{(i,j)}) = 0 \forall (i, j) \neq (1, 0)$

If there exists a direction $\vec{v}$ such that $h(\vec{v}) > 0$, then the growth rate of orbits along R_n is at least in the order of 2^n . Directional entropy can be regarded as a generalization of the entropy of

noncocompact subgroup action of $\mathbb{Z}^2$. To understand entropy zero $\mathbb{Z}^d$ -actions ($d > 2$), we need to investigate the complexity of noncocompact subgroup actions of $\mathbb{Z}^i$ ($0 \leq i \leq d-1$) in addition to directions.

In the case of $\mathbb{Z}^2$ -action generated by a shift and a Cellular Automaton map (a block code), J. Milnor asked if the directional entropy function is continuous. We have the following answer (Park 1999).

Theorem 27 *Directional entropy $h(\vec{v})$ is continuous on $\|\vec{v}\| = 1$ for the action σ generated by a shift T and a Cellular Automaton map ϕ .*

Sinai (2017) has shown that $h(\vec{v})$ is upper semi-continuous on $\|\vec{v}\| = 1$. In fact, the continuity is guaranteed for more general σ . Topological directional entropy $h_{\text{top}}(\vec{v})$ can be defined analogously. It is not yet known if $h_{\text{top}}(\vec{v})$ is continuous in this restrictive class of $\mathbb{Z}^2$ -actions.

Let $(X, \mathcal{X}, \mu, \sigma)$ be a subshift of $\{0, 1\}^{\mathbb{Z}^2}$. As was remarked via the examples mentioned above, entropy zero systems exhibit diverse behavior. For the growth rate of orbits of zero entropy systems, several notions, slow entropy (Katok and Thouvenot 1997) and entropy dimension (Dou et al. 2019; Jung et al. 2017), have been introduced to understand the complexity. They intend to measure the intermediate growth rate which is bigger than polynomial and less than exponential. Although the motivation is mainly from the study of bigger group actions, investigation of the growth rate of orbits in entropy zero $\mathbb{Z}$ -actions is necessary for more complete understanding.

Finitely Observable Invariant

Having given the basic properties of entropy, we would like to make a remark on its characteristic of finitely observable invariant in ergodic systems. In a landmark paper (Ornstein and Weiss 2007), Ornstein and Weiss show that all “finitely observable” properties of ergodic processes are secretly entropy; indeed, they are continuous

functions of entropy. This was generalized by Gutman and Hochman (2008); some of the notation below is from their paper.

Here is the setting. Consider an ergodic process, on a nonatomic space, taking on only finitely many values in $\mathbb{N}$; let $\mathcal{C}$ be some family of such processes. An **observation scheme** is a metric space (Ω, d) and a sequence of functions $\mathbf{S} = (S_n)_1^\infty$, where S_n maps $\mathbb{N} \times \dots \times \mathbb{N}$ into Ω . On a point $\vec{x} \in \mathbb{N}^\infty$, the scheme **converges** if

$$n \mapsto S_n(x_1, x_2, \dots, x_n) \quad (35)$$

converges in Ω . And on a particular process $\mathcal{X}$, say that $\mathbf{S}$ **converges**, if $\mathbf{S}$ converges on a.e. $\vec{x}$ in $\mathcal{X}$.

A function $J : \mathcal{C} \rightarrow \Omega$ is isomorphism invariant if, whenever the underlying transformations of two processes $\mathcal{X}, \mathcal{X}' \in \mathcal{C}$ are isomorphic, then $J(\mathcal{X}) = J(\mathcal{X}')$. Lastly, say that $\mathbf{S}$ “converges to J ,” if for each $\mathcal{X} \in \mathcal{C}$, scheme $\mathbf{S}$ converges to the value $J(\mathcal{X})$.

The work of David Bailey (1976) produced an observation scheme for entropy. The Lempel-Ziv algorithm (Ziv and Lempel 1977) was another entropy observer, with practical application (Shields 1996)

Ornstein and Weiss provided entropy schemes in (Ornstein and Weiss 1993; Ornstein et al. 1990). Their recent paper “*Entropy is the only finitely-observable invariant*” (Ornstein and Weiss 2007) gives a converse, a uniqueness result.

Theorem 28 (Ornstein-Weiss) *Suppose J is a finitely observable function, defined on all ergodic finite-valued processes. If J is an isomorphism invariant, then J is a continuous function of the entropy.*

Finitely Observable Extension

Extending the Ornstein-Weiss result, Yonatan Gutman and Michael Hochman, in Gutman and Hochman (2008) proved that it holds even when the isomorphism invariant, J , is well defined only on certain subclasses of the set of all ergodic processes. In particular they obtain the following result on three classes of zero-entropy transformations.

Theorem 29 (Gutman-Hochman) *Suppose $J()$ is a finitely observable invariant on one of the following classes:*

- (i) *The Kronecker systems; the class of systems with pure point spectrum.*
- (ii) *The zero-entropy mild mixing processes.*
- (iii) *The zero-entropy strong mixing processes.*

Then $J()$ is constant.

Recent Progress

In addition to a survey of older results in measure-theoretic entropy and in topological entropy, let us end this entry with a brief discussion of a couple of recent results, chosen from many.

Entropy of Actions of Nonamenable Groups

Consider $(G, \mathcal{G})$, a topological group and its Borel field (σ -algebra). Let $(\mathcal{G} \times \mathcal{X})$ be the field on $G \times X$ generated by the two coordinate-subfields. A map

$$\psi : G \times X \rightarrow X \quad (36)$$

is **measurable** if

$$\psi^{-1}(x) \subset \mathcal{G} \times X.$$

Use $\psi^g(x)$ for $\psi(g, x)$.

This map in (36) is a (measure-preserving) **group action** if $\forall g, h \in G : \psi^g \circ \psi^h = \psi^{gh}$, and each $\psi^g : X \rightarrow X$ is measure-preserving.

This encyclopedia entry has mainly discussed entropy for $\mathbb{Z}$ -actions, that is, when $G = \mathbb{Z}$. The ergodic theorem, our definition of entropy, and large parts of ergodic theory involve taking averages (of some quantity of interest) over larger and larger “pieces of time.” In $\mathbb{Z}$ or $\mathbb{R}$, we typically use the intervals $I_k := [0 \dots k) \subset \mathbb{Z}$ or $I_k := [0 \dots k) \subset \mathbb{R}$. When G is $\mathbb{Z}^d$, we average over squares $R_k = \underbrace{I_k \times I_k \times \dots \times I_k}_{d\text{-times}}$.

The **amenable groups** are those which possess, in a certain sense, larger and larger averaging sets

called Følner sequence. Parts of ergodic theory have been carried over to actions of amenable groups, for example, Ornstein and Weiss (1983) and Ward and Zhang (1992). Indeed, much of the entropy theory was extended to certain amenable groups by Ornstein and Weiss (1987). It is remarkable that many notions established for $\mathbb{Z}$ -actions played an essential role in developing the properties of amenable group actions, especially the notion of orbit equivalence (Connes et al. 1981; Danilenko and Park 2002; Rudolph and Weiss 2000).

The typical example of a *nonamenable* group is a free group (on more than one generator). Pursuit to generalize the entropy theory to non-amenable groups had been halted by the simple, but unexpected example by Ornstein and Weiss: Let $\mathbb{F}_2$ be a free group of two generators. There exists a factor map from the Bernoulli action of distribution $\{\frac{1}{2}, \frac{1}{2}\}$ to the Bernoulli action of distribution $\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$. That is, a factor has a bigger entropy in the case of $\mathbb{F}_2$ -action, unlike amenable group actions. Bowen has opened the door to the world of entropy theory of non-amenable group actions. Bowen (2010) succeeded in extending the definition of entropy to actions of finite-rank free groups.

Theorem 30 (L. Bowen) *Let G be a finite-rank free group. Then two Bernoulli G -actions are isomorphic IFF they have the same entropy.*

The paper introduces an isomorphism invariant, the “ f invariant,” and shows that, for Bernoulli actions, the f invariant agrees with entropy, that is, with the distropy of the independent generating partition.

And it was not too long to introduce a new invariant, called sofic entropy to extend the result to a much larger class of groups. The sofic entropy is defined via a sofic approximation of a group (Bowen 2012, 2017).

Let G be a countable discrete group. We say G is sofic if there exists $\sigma_n : G \rightarrow \text{Sym}(\{0, 1, \dots, m_n - 1\})$ into permutation groups satisfying

- (i) $\lim_{n \rightarrow \infty} \frac{1}{m_n} \#\{k \in \{0, 1, \dots, m_n - 1\} : \sigma_n(gh)k = \sigma_n(g)\sigma_n(h)k\} = 1 \text{ for all } g, h \in G.$
- (ii) $\lim_{n \rightarrow \infty} \frac{1}{m_n} \#\{k \in \{0, 1, \dots, m_n - 1\} : \sigma_n(g)k \neq k\} = 1 \text{ for all } g \in G \setminus \{1\}.$

We call $\Sigma = \{\sigma_n\}$ a sofic approximation. Properties (i) and (ii) say that the sequence is asymptotically homomorphism and asymptotically free⁴.

Theorem 31 *Let G be a sofic group and let $(X, \mathcal{X}, \mu, \sigma)$ be a dynamical system of G -action. Two Bernoulli actions are isomorphic IFF their distribution entropies of independent partitions are the same.*

He extended the above theorem to countable groups.

Theorem 32 *Let G be a countable nonamenable group. If two partitions have the same distribution entropy with at least 3-states, then two Bernoulli actions of G are isomorphic.*

B. Seward (2018) was able to eliminate “at least 3-states” condition. If G is amenable, the sofic entropy agrees with classical entropy. The sofic, measure and topological, entropies are known to be dependent on the sofic approximations, unlike amenable groups (Bowen 2017, Bowen). However, for Bernoulli shifts of G the sofic entropy is independent of the sofic approximation.

D. Kerr and H. Li (2011) introduced topological sofic entropy on a compact metric space meaning the exponential growth rate of the number of “approximate periodic points (pseudo-periodic points).” This leads them to prove the following variational principle.

⁴The notion of sofic group is due to Mikheal Gromov (1999). There is yet no known example of a countable group which is not sofic.

Theorem 33 (Kerr-Li) *Let G be a countable nonamenable group action on (X, σ) . For a sofic approximate Σ of G ,*

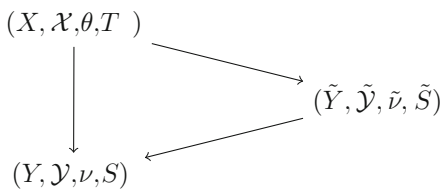
$$h_{\Sigma, \text{top}}(\sigma) = \sup_{\mu} h_{\Sigma}(\sigma, \mu)$$

where μ is an invariant measure on (X, σ) .

Weak Pinsker Conjecture

Pinsker conjectured that every measure-preserving transformation is a product of K -automorphism and 0-entropy transformation (Pinsker 1960). Pesin (1977) showed that some flows on surfaces have the property; Bernoulli $\times$ 0-entropy (in fact, a rotation). However, by Ornstein's counterexample (Ornstein 1973) to the conjecture, it is shown to be false in general. Thouvenot (1977) proposed a weaker version of the conjecture, called weak Pinsker conjecture that every measure-preserving transformation is relatively Bernoulli over a transformation of arbitrarily small entropy. That is, given any $\epsilon > 0$, $(X, \mathcal{X}, \mu, T)$ is isomorphic to a product Bernoulli $\times$ (transformation of entropy $< \epsilon$). Austin (2018) proves the conjecture is the affirmative. In fact, he proves a little more.

Theorem 34 (Austin) *If $(X, \mathcal{X}, \mu, T)$ has a factor $(Y, \mathcal{Y}, \nu, S)$ and $h(T) > h(S)$ then for a given $\epsilon > 0$, there exists an extension $(\tilde{Y}, \tilde{\mathcal{Y}}, \tilde{\nu}, \tilde{S})$ of $(Y, \mathcal{Y}, \nu, S)$ such that the following diagram holds:*



where

- (i) $h(\tilde{S}) < h(S) + \epsilon$.
- (ii) $(X, \mathcal{X}, \mu, T)$ is isomorphic to Bernoulli $\times (\tilde{Y}, \tilde{\mathcal{Y}}, \tilde{\nu}, \tilde{S})$.

Exodos

Ever since the pioneering work of Shannon, and of Kolmogorov and Sinai, entropy has been front and

center as a major tool in Ergodic Theory and Topological Dynamics. Simply mentioning all the substantial results in entropy theory would dwarf the length of this encyclopedia entry many times over. And as the recent results mentioned above (cherry-picked out of many) show, Entropy shows no sign of fading away. . .

Bibliography

- Adler RL, Marcus B (1979) Topological entropy and equivalence of dynamical systems. *Memoirs of the American Mathematical Society*, vol 219. American Mathematical Society, Providence
- Adler RL, Weiss B (1967) Entropy, a complete metric invariant for automorphisms of the torus. *Proc Natl Acad Sci USA* 57(6):1573
- Adler RL, Konheim AG, McAndrew MH (1965) Topological entropy. *Trans Am Math Soc* 114(2):309–319
- Ambrose W (1941) Representation of ergodic flows. *Ann Math* 42:723–739
- Austin T (2014) Scenery entropy as an invariant of RWRS processes. *arXiv preprint arXiv:1405.1468*
- Austin T (2018) Measure concentration and the weak pinsker property. *Publ Math IHÉS* 128(1):1–119
- Bailey DH (1976) Sequential schemes for classifying and predicting ergodic processes. PhD thesis, Stanford University
- Berg KR (1975) Independence and additive entropy. *Proc Am Math Soc* 51(2):366–370
- Bowen R (1971) Entropy for group endomorphisms and homogeneous spaces. *Trans Am Math Soc* 153:401–414
- Bowen R (1973) Topological entropy for noncompact sets. *Trans Am Math Soc* 184:125–136
- Bowen L (2010) A new measure conjugacy invariant for actions of free groups. *Ann Math* 171(2):1387–1400
- Bowen L (2012) Every countably infinite group is almost Ornstein. In: *Dynamical systems and group actions*, vol 567. American Mathematical Society, Providence, pp 67–78
- Bowen L (2017) Examples in the entropy theory of countable group actions. *arXiv:1704.06349 [math.DS]*, pp 1–131
- Brin M, Stuck G (2002) *Introduction to dynamical systems*. Cambridge University Press, Cambridge
- Buzzi J, Ruelle S (2006) Large entropy implies existence of a maximal entropy measure for interval maps. *Discrete Contin Dynam Syst A* 14(4):673–688
- Clausius R (1864) *Abhandlungen über die mechanische Wärmetheorie*, vol 1. F. Vieweg, Braunschweig
- Clausius R (1867) *Abhandlungen über die mechanische Wärmetheorie*, vol 2. F. Vieweg, Braunschweig
- Connes A, Feldman J, Weiss B (1981) An amenable equivalence relation is generated by a single transformation. *Ergodic Theory Dynam Syst* 1(4):431–450
- Cornfeld I, Fomin S, Sinai YG (1982) *Ergodic theory*. Grundlehren der mathematischen Wissenschaften: a

- series of comprehensive studies in mathematics. Springer, New York
- Danilenko AI, Park KK (2002) Generators and bernoullian factors for amenable actions and cocycles on their orbits. *Ergodic Theory Dynam Syst* 22(6):1715–1745
- den Hollander F, Steif JE et al (2006) Random walk in random scenery: a survey of some recent results. In: *Dynamics & stochastics*. Institute of mathematical Statistics lecture Notes 48:53–65
- Denker M (1976) Measures with maximal entropy. In: *Théorie Ergodique*. Springer, Berlin, pp 70–112
- Dinaburg EI (1970) A correlation between topological entropy and metric entropy. *Dokl Akad Nauk* 190:19–22. Russian Academy of Sciences
- Dou D, Huang W, Park KK (2019) Entropy dimension of measure preserving systems. *Trans Am Math Soc* 371(10):7029–7065
- Ferenczi S (1997) Systems of finite rank. *Colloq Math* 73:35–65
- Friedman N (1970) Introduction to ergodic theory. Van Nostrand Reinhold mathematical studies. Van Nostrand Reinhold, New York
- Gromov M (1999) Endomorphisms of symbolic algebraic varieties. *J Eur Math Soc* 1(2):109–197
- Gutman Y, Hochman M (2008) On processes which cannot be distinguished by finite observation. *Isr J Math* 164(1):265–284
- Halmos P (1956) *Lectures on ergodic theory*. Chelsea scientific books. Mathematical Society of Japan, Tokyo
- Hasselblatt B, Nitecki Z, Propp J (2005) Topological entropy for non-uniformly continuous maps. *arXiv preprint math/0511495*
- Jung U, Lee J, Park KK (2017) Topological entropy dimension and directional entropy dimension for $\mathbb{Z}^2$ -subshifts. *Entropy* 19(2):46
- Kalikow SA (1982) T, t-1 transformation is not loosely Bernoulli. *Ann Math* 115:393–409
- Kalikow S, King JL (1994) A countably-valued sleeping stockbroker process. *J Theor Probab* 7(4):703–708
- Katok A (1980) Smooth non-Bernoulli K-automorphisms. *Invent Math* 61(3):291–299
- Katok A, Hasselblatt B (1995) Introduction to the modern theory of dynamical systems. In: *Encyclopedia of mathematics and its applications*. Cambridge University Press, Cambridge
- Katok A, Thouvenot J-P (1997) Slow entropy type invariants and smooth realization of commuting measure-preserving transformations. *Ann Inst Henri Poincaré Probab Stat* 33:323–338. Elsevier
- Kerr D, Li H (2011) Entropy and the variational principle for actions of sofic groups. *Invent Math* 186(3):501–558
- King JL (1988) Joining-rank and the structure of finite rank mixing transformations. *J Anal Math* 51(1):182–227
- King JL (1992) Dilemma of the sleeping stockbroker. *Am Math Mon* 99(4):335–338
- Kolmogorov A (1958) A new metric invariant of transitive dynamical systems and lebesgue space automorphisms. *Dokl Akad Sci USSR* 119:861–864
- Krieger W (1970) On entropy and generators of measurepreserving transformations. *Trans Am Math Soc* 149(2):453–464
- Ledrappier F, Young L-S (1985) The metric entropy of diffeomorphisms: Part I: characterization of measures satisfying Pesin's entropy formula. *Ann Math* 122:509–574
- Lind D, Marcus B (1996) *An introduction to symbolic dynamics and coding*. Cambridge University Press, Cambridge
- Mane R (1987) *Ergodic theory and differentiable dynamics*. Springer, Berlin/Heidelberg
- Meshalkin L (1959) A case of isomorphism of bernoulli schemes. *Dokl Akad Nauk SSSR* 128(1):41–44
- Milnor J (1988) On the entropy geometry of cellular automata. *Complex Syst* 2(3):357–385
- Misiurewicz M (1973) Diffeomorphism without any measure with maximal entropy. *Bull Acad Pol Sci Sér Sci Math Astron Phys* 21(10):903–910
- Ornstein DS (1970) Bernoulli shifts with the same entropy are isomorphic. *Adv Math* 4(3):337–352
- Ornstein DS (1973) A mixing transformation for which Pinsker's conjecture fails. *Adv Math* 10(1):103–123
- Ornstein DS (1974) *Ergodic theory, randomness and dynamical systems*, no 5. Yale University Press, New Haven
- Ornstein DS (2013) Newton's laws and coin tossing. *Not Am Math Soc* 60(4):450
- Ornstein DS, Shields PC (1973) An uncountable family of k-automorphisms. *Adv Math* 10(1):63–88
- Ornstein DS, Weiss B (1974) Finitely determined implies very weak Bernoulli. *Isr J Math* 17(1):94–104
- Ornstein DS, Weiss B (1983) The Shannon-McMillan-Breiman theorem for a class of amenable groups. *Isr J Math* 44(1):53–60
- Ornstein DS, Weiss B (1987) Entropy and isomorphism theorems for actions of amenable groups. *J Anal Math* 48(1):1–141
- Ornstein DS, Weiss B (1993) Entropy and data compression schemes. *IEEE Trans Inf Theory* 39(1):78–83
- Ornstein DS, Weiss B (2007) Entropy is the only finitely-observable invariant. *J Mod Dyn* 1(1):93–105
- Ornstein DS, Weiss B et al (1990) How sampling reveals a process. *Ann Probab* 18(3):905–930
- Park KK (1999) On directional entropy functions. *Isr J Math* 113(1):243–267
- Pesin YB (1977) Characteristic Lyapunov exponents and smooth ergodic theory. *Usp Matematicheskikh Nauk* 32(4):55–112
- Petersen K (1983) *Ergodic theory*. Cambridge studies in advanced mathematics. Cambridge University Press, Cambridge
- Pinsker MS (1960) Dynamical systems with completely positive or zero entropy. *Dokl Akad Nauk* 133:1025–1026. Russian Academy of Sciences
- Rudin W (1973) *Functional analysis*. McGraw-Hill series in higher mathematics. McGraw-Hill, New York
- Rudolph D (1976) A two-valued step coding for ergodic flows. *Math Z* 150(3):201–220

- Rudolph D (1990) Fundamentals of measurable dynamics: ergodic theory on Lebesgue spaces. Oxford science publications. Clarendon Press, Oxford
- Rudolph DJ, Weiss B (2000) Entropy and mixing for amenable group actions. *Ann Math* 151(3):1119–1150
- Seward B (2018) Bernoulli shifts with bases of equal entropy are isomorphic. *arXiv preprint arXiv:1805.08279*
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423
- Shields PC (1973) The theory of Bernoulli shifts. University of Chicago Press, Chicago
- Shields PC (1996) The ergodic theory of discrete sample paths, vol 13. American Mathematical Society, Providence
- Sinai YG (1959) On the concept of entropy for a dynamic system. *Dokl Akad Nauk SSSR* 124(4):768–771
- Sinai YG (1962) A weak isomorphism of transformations with invariant measure. *Dokl Akad Nauk* 147(4):177–207
- Sinai YG (2017) Topics in ergodic theory, vol 44. Princeton University Press, Princeton
- Thouvenot J-P (1975) Quelques propriétés des systèmes dynamiques qui se décomposent en un produit de deux systèmes dont l'un est un schéma de Bernoulli. *Isr J Math* 21(2–3):177–207
- Thouvenot J-P (1977) On the stability of the weak Pinsker property. *Isr J Math* 27(2):150–162
- Tribus M, McIrvine EC (1971) Energy and information. *Sci Am* 225(3):179–188
- Walters P (1981) An introduction to ergodic theory. Graduate texts in mathematics. Springer, New York
- Ward T, Zhang Q (1992) The Abramov-Rokhlin entropy addition formula for amenable group actions. *Monatsh Math* 114(3):317–329
- Young L-S (1982) Dimension, entropy and Lyapunov exponents. *Ergodic Theory Dynam Syst* 2(1):109–124
- Ziv J, Lempel A (1977) A universal algorithm for sequential data compression. *IEEE Trans Inf Theory* 23(3):337–343

Articles, Papers, and Books of Interest

- Bowen L. <https://warwick.ac.uk/fac/sci/math/research/events/2018-19/metc/abstracts>
- Boyle M, Downarowicz T (2004) The entropy theory of symbolic extensions. *Invent Math* 156(1):119–161
- Downarowicz T (2011) Entropy in dynamical systems. Cambridge University Press, Cambridge
- Downarowicz T, Serafin J (2003) Possible entropy functions. *Isr J Math* 135(1):221–250
- Greven A, Keller G, Warnecke G (2003) Entropy. Princeton series in applied mathematics. Princeton University Press, Princeton
- Hassner M (1980) A non-probabilistic source and channel coding theory. PhD thesis, UCLA
- Katok A (2007) Fifty years of entropy in dynamics: 1958–2007. *J Modern Dyn* 1(4):545–596
- Katok A, Sinai YG, Stepin AM (1977) Theory of dynamical systems and general transformation groups with invariant measure. *J Sov Math* 7(6):974–1065
- McMillan B (1953) The basic theorems of information theory. *Ann Math Statist* 24(2):196–219
- Newhouse SE (1989) Continuity properties of entropy. *Ann Math* 129(1):215–235
- Parry W (1969) Entropy and generators in ergodic theory. Mathematics lecture note series. W. A. Benjamin, New York
- Wikipedia. http://en.wikipedia.org/wiki/Spectral_radius, http://en.wikipedia.org/wiki/Information_entropy

Isomorphism Theory in Ergodic Theory

Christopher Hoffman

Department of Mathematics, University of Washington, Seattle, WA, USA

Article Outline

Glossary

Definition of the Subject

Introduction

Basic Transformations

Basic Isomorphism Invariants

Basic Tools

Isomorphism of Bernoulli Shifts

Transformations Isomorphic to Bernoulli Shifts

Transformations Not Isomorphic to Bernoulli Shifts

Classifying the Invariant Measures of Algebraic Actions

Finitary Isomorphisms

Flows

Other Equivalence Relations

Non-invertible Transformations

Factors of a Transformation

Actions of Amenable Groups

Future Directions

Bibliography

Glossary

Almost everywhere A property is said to hold **almost everywhere (a.e.)** if the set on which the property does not hold has measure 0.

Bernoulli shift A Bernoulli shift is a stochastic process such that all outputs of the process are independent.

Conditional measure For any measure space $(X, \mathcal{B}, \mu)$ and σ -algebra $\mathcal{C} \subset \mathcal{B}$ the conditional measure is a $\mathcal{C}$ -measurable function g such that $\mu(C) = \int_C g \, d\mu$ for all $C \in \mathcal{C}$.

Coupling of two measure spaces A coupling of two measure spaces $(X, \mu, \mathcal{B})$ and $(Y, \nu, \mathcal{C})$ is a measure γ on $X \times Y$ such that $\gamma(B \times Y) = \mu(B)$ for all $B \in \mathcal{B}$ and $\gamma(X \times C) = \nu(C)$ for all $C \in \mathcal{C}$.

Ergodic measure preserving transformation A measure preserving transformation is ergodic if the only invariant sets $(\mu(A \Delta T^{-1}(A)) = 0)$ have measure 0 or 1.

Ergodic theorem The pointwise ergodic theorem says that for any measure preserving transformation $(X, \mathcal{B}, \mu)$ and T and any L^1 function f the time average $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(T^i(x))$ converges

a.e. If the transformation is ergodic then the limit is the space average, $\int f \, d\mu$ a.e.

Geodesic A geodesic on a Riemannian manifold is a distance minimizing path between points.

Horocycle A horocycle is a circle in the hyperbolic disk which intersects the boundary of the disk in exactly one point.

Invariant measure Likewise a measure μ is said to be invariant with respect to $(\mathbf{X}, T)$ provided that $\mu(T^{-1}(A)) = \mu(A)$ for all measurable $A \in \mathcal{B}$.

Joining of two measure preserving transformations A joining of two measure preserving transformations $(\mathbf{X}, T)$ and $(\mathbf{Y}, S)$ is a coupling of $\mathbf{X}$ and $\mathbf{Y}$ which is invariant under $T \times S$.

Markov shift A Markov shift is a stochastic process such that the conditional distribution of the future outputs $(\{x_n\}_{n>0})$ of the process conditioned on the last output (x_0) is the same as the distribution conditioned on all of the past outputs of the process $(\{x_n\}_{n \leq 0})$.

Measure preserving transformation A measure preserving transformation consists of a probability space $(\mathbf{X}, T)$ and a measurable function $T : X \rightarrow X$ such that $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}$.

Measure theoretic entropy A numerical invariant of measure preserving transformations that measures the growth in complexity of

measurable partitions refined under the iteration of the transformation.

Probability space A probability space $\mathbf{X} = (X, \mu, \mathcal{B})$ is a measure space such that $\mu(B) = 1$.

Rational function, rational map A rational function $f(z) = g(z)/h(z)$ is the quotient of two polynomials. The degree of $f(z)$ is the maximum of the degrees of $g(z)$ and $h(z)$. The corresponding rational maps $T_f: z \rightarrow f(z)$ on the Riemann sphere $\mathbb{C}$ are a main object of study in complex dynamics.

Stochastic process A stochastic process is a sequence of measurable functions $\{x_n\}_{n \in \mathbb{Z}}$ (or outputs) defined on the same measure space, $\mathbf{X}$. We refer to the value of the functions as outputs.

Definition of the Subject

Our main goal in this article is to consider when two measure preserving transformations are in some sense different presentations of the same underlying object. To make this precise we say two measure preserving maps $(\mathbf{X}, T)$ and $(\mathbf{Y}, S)$ are **isomorphic** if there exists a measurable map $\phi: X \rightarrow Y$ such that

1. ϕ is measure preserving,
2. ϕ is invertible almost everywhere and
3. $\phi(T(x)) = S(\phi(x))$ for almost every x .

The main goal of the subject is to construct a collection of invariants of a transformation such that a necessary condition for two transformations to be isomorphic is that the invariant be the same for both transformations. Another goal of the subject is to solve the much more difficult problem of constructing invariants such that the invariants being the same for two transformations is a sufficient condition for the transformation to be isomorphic. Finally we apply these invariants to many natural classes of transformation to see which of them are (or are not) isomorphic.

Introduction

In this article we look at the problem of determining of which measure preserving transformations are isomorphic. We look at a number of isomorphism invariants, the most important of which is the Kolmogorov–Sinai entropy. The central theorem in this field is Ornstein’s proof that any two Bernoulli shifts of the same entropy are isomorphic. We also discuss some of the consequences of this theorem, which transformations are isomorphic to Bernoulli shifts as well as generalizations of Ornstein’s theory.

Basic Transformations

In this section we list some of the basic classes of measure preserving transformations that we study in this article.

Bernoulli shifts Some of the most fundamental transformations are the Bernoulli shifts. A **probability vector** is a vector $\{p_i\}_{i=1}^n$ such that $\sum_{i=1}^n p_i = 1$ and $p_i \geq 0$ for all i . Let $\mathbf{p} = \{p_i\}_{i=1}^n$ be a probability vector. The **Bernoulli shift** corresponding to $\mathbf{p}$ has state space $\{1, 2, \dots, n\}^{\mathbb{Z}}$, the shift operator $T(x)_i = x_{i+1}$. To specify the measure we only need to specify it on **cylinder sets**

$$A = \{x \in X: x_i = a_i \ \forall i \in \{m, \dots, k\}\}$$

for some $m \leq k \in \mathbb{Z}$ and a sequence $a_m, \dots, a_k \in \{1, \dots, n\}$. The measure on cylinder sets is defined by

$$\mu\{x \in X: x_i = a_i \ [-2ex] \text{ for all } i \text{ such that } m \leq i \leq k\} \\ = \prod_{i=m}^k p_{a_i}.$$

For any $d \in \mathbb{N}$ if $\mathbf{p} = (1/d, \dots, 1/d)$ we refer to $\text{Bernoulli}_{\mathbf{p}}$ as the **Bernoulli d shift**.

Markov shifts A **Markov shift** on state n symbols is defined by an $n \times n$ matrix, M , such that $\{M(i, j)\}_{j=1}^n$ is a probability vector for each i . The Markov shift is a measure preserving

transformation with state space $\{1, 2, \dots, n\}^{\mathbb{Z}}$, transformation $T(x)_i = x_{i+1}$

$$\begin{aligned} & \mu\{x_1 = a_1 \mid x_0 = a_0\} \\ &= \mu\{x_1 = a_1 \mid x_0 = a_0, x_{-1} = a_{-1}, x_{-2} = a_{-2}, \dots\} \end{aligned}$$

for all choices of $a_i, i \leq 1$. Let $m = \{m(i)\}_{i=1}^n$ be a vector such that $Mm = m$. Then an invariant measure is defined by setting the measure on cylinder sets to be $A = \{x : x_0 = a_0, x_1 = a_1, \dots, x_n = a_n\}$ is given by $\mu(A) = m(a_0) \prod_{i=1}^n M(a_{i-1}, a_i)$.

Shift maps More generally the **shift map** σ is the map $\sigma : \mathbb{N}^{\mathbb{Z}} \rightarrow \mathbb{N}^{\mathbb{Z}}$ where $\sigma(x)_i = x_{i+1}$ for all $x \in \mathbb{N}^{\mathbb{Z}}$ and $i \in \mathbb{Z}$. We also let σ designate the shift map on $\mathbb{N}^{\mathbb{N}}$. For each measure that is invariant under the shift map there is a corresponding measure defined on $\mathbb{N}^{\mathbb{N}}$ that is invariant under the shift map. Let μ be an invariant measure under the shift map. For any measurable set of $A \subset \mathbb{N}^{\mathbb{N}}$ we define $\tilde{A}$ on $\mathbb{N}^{\mathbb{Z}}$ by

$$\tilde{A} = \{\dots, x_{-1}, x_0, x_1, \dots : x_0, x_1, \dots \in A\}.$$

Then it is easy to check that $\tilde{\mu}$ defined by $\tilde{\mu}(\tilde{A}) = \mu(A)$ is invariant. If the original transformation was a Markov or Bernoulli shift then refer to the resulting transformations as a **one sided Markov shifts** or **one sided Bernoulli shift** respectively.

Rational maps of the Riemann sphere We say that $f(z) = g(z)/h(z)$ is a **rational function of degree** $d \geq 2$ if both $g(z)$ and $h(z)$ are polynomials with $\max(\deg(g(z)), \deg(h(z))) = d$. Then f induces a natural action on the Riemann sphere $T_f : z \rightarrow f(z)$ which is a d to one map (counting with multiplicity). In section “Rational Maps” we shall see that for every rational function f there is a canonical measure μ_f such that T_f is a measure preserving transformation.

Horocycle flows The horocycle flow acts on $SL(2, \mathbb{R})/\Gamma$ where Γ is a discrete subgroup of $SL(2, \mathbb{R})$ such that $SL(2, \mathbb{R})/\Gamma$ has finite Haar measure. For any $g \in SL(2, \mathbb{R})$ and $t \in \mathbb{R}$ we define the horocycle flow by

$$h_t(\Gamma g) = \Gamma g \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}.$$

Matrix actions Another natural class of actions is given by the action of matrices on tori. Let M be an invertible $n \times n$ integer valued matrix. We define $T_M : [0, 1)^n \rightarrow [0, 1)^n$ by $T_M(x)_i = M(x)_i \bmod 1$ for all $i, 1 \leq i \leq n$. It is easy to check that if M is surjective then Lebesgue measure is invariant under T_M . T_M is a $|\det(M)|$ to one map. If $n = 1$ then M is an integer and we refer to the map as times M .

The $[T, T^{-1}]$ transformations Let (X, T) be any invertible measure preserving transformation. Let σ be the shift operator on $Y = \{-1, 1\}^{\mathbb{Z}}$. The space Y comes equipped with the Bernoulli $(1/2, 1/2)$ product measure ν . The $[T, T^{-1}]$ transformation is a map on $Y \times X$ which preserves $\nu \times \mu$. It is defined by

$$T, T^{-1}(y, x) = (S(y), T^{y_0}(x)).$$

Induced transformations Let (X, T) be a measure preserving transformation and let $A \subset X$ with $0 < \mu(A) < 1$. The **transformation induced by** $A, (A, T_A, \mu_A)$, is defined as follows. For any $x \in A$

$$T_A(x) = T^{n(x)}(x)$$

where $n(x) = \inf \{m > 0 : T^m(x) \in A\}$. For any $B \subset A$ we have that $\mu_A(B) = \mu(B)/\mu(A)$.

Basic Isomorphism Invariants

The main purpose of isomorphism theory is to classify which pairs of measure preserving transformation are isomorphic and which are not isomorphic. One of the main ways that we can show that two measure preserving transformation are not isomorphic is using isomorphism invariants. An **isomorphism invariant** is a function f defined on measure preserving transformations such that if (X, T) is isomorphic to (Y, S) then $f((X, T)) = f((Y, S))$.

A measure preserving action is said to be **ergodic** if $\mu(A) = 0$ or 1 for every A with

$$\mu(A \Delta T^{-1}(A)) = 0.$$

A measure preserving action is said to be **weak mixing** if for every measurable A and B

$$\frac{1}{n} \sum_{n=1}^{\infty} |\mu(A \cap T^{-n}(B)) - \mu(A)\mu(B)| = 0.$$

A measure preserving action is said to be **mixing** if for every measurable A and B

$$\lim_{n \rightarrow \infty} \mu(A \cap T^{-n}(B)) = \mu(A)\mu(B).$$

It is easy to show that all three of these properties are isomorphism invariants and that

$$\begin{aligned} (\mathbf{X}, T) \text{ is mixing} &\Rightarrow (\mathbf{X}, T) \text{ is weak mixing} \\ &\Rightarrow (\mathbf{X}, T) \text{ is ergodic.} \end{aligned}$$

We include one more definition before introducing an even stronger isomorphism invariant. The **action of a group** (or a semigroup) G on a probability space $(X, \mu, \mathcal{B})$ is a family of measure preserving transformations $\{f_g\}_{g \in G}$ such that for any $g, h \in G$ we have that $f_g(f_h(x)) = f_{g+h}(x)$ for almost every $x \in X$. Thus any invertible measure preserving transformation $(\mathbf{X}, T)$ induces a $\mathbb{Z}$ action by $f_n(x) = T^n(x)$.

We say that a group action $(\mathbf{X}, T_g)$ is mixing of all orders if for every $n \in \mathbb{N}$ and every collection of set $A_1, \dots, A_n \in \mathcal{B}$ then

$$\begin{aligned} \mu(A_1 \cap T_{g_2}(A_2) \cap T_{g_2+g_3}(A_3) \dots \\ \cap T_{g_2+g_3+\dots+g_n}(A_n)) \rightarrow \prod_{i=1}^n \mu(A_i) \end{aligned}$$

as the g_i go to infinity.

Basic Tools

Birkhoff's ergodic theorem states that for every measure preserving action the limits

$$\widehat{f}(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)$$

exists for almost every x and if $(\mathbf{X}, T)$ is ergodic then the limit is $\bar{f} = \int f d\mu$ (Birkhoff 1931).

A **partition** P of X is a measurable function defined on X . (For simplicity we often assume that a partition is a function to $\mathbb{Z}$ or some subset of $\mathbb{Z}$.) We write P_i for $P^{-1}(i)$. For any partition P of $(\mathbf{X}, T)$ define the partition $T^i P$ by $P \circ T^i$. Thus for invertible T this is given by $T^i P(x) = P(T^{-i}(x))$. Then define $(P)_T = \bigvee_{i \in \mathbb{Z}} T^i P$. Thus $(P)_T$ is the smallest σ -algebra which contains $T^i(P_j)$ for all $i \in \mathbb{Z}$ and $j \in \mathbb{N}$. We say that $(P)_T$ is the **σ -algebra generated by P** . A partition P is a **generator** of $(\mathbf{X}, T)$ if $(P)_T = \mathcal{B}$. Many measure preserving transformations come equipped with a natural partition.

Rokhlin's theorem For any measure preserving transformation $(\mathbf{X}, T)$, any $\epsilon > 0$ and any $n \in \mathbb{N}$ there exists $A \subset X$ such that $T^{-i}(A) \cap T^{-j}(A)$ for all $0 \leq i < j \leq n$ and $\mu(\bigcup_{i=0}^n T^i(A)) > 1 - \epsilon$. Moreover for any finite partition P of X we can choose A such that $\mu(P_i \cap A) = \mu(A)\mu(P_i)$ for all $i \in \mathbb{N}$ (Halmos 1960).

Shannon–McMillan–Breiman theorem (Breiman 1957; Petersen 1989) For any measure preserving system $(\mathbf{X}, T)$ and any $\epsilon > 0$ there exists $n \in \mathbb{N}$ and a set G with $\mu(G) > 1 - \epsilon$ with the following property. For any sequence $g_1, \dots, g_n \in \mathbb{N}$ let $g = \bigcap_{i=1}^n P(T^{-i}(X)) = g_i$. Then if $\mu(g \cap G) > 313\epsilon^2$ then

$$\mu(g) \in \left(2^{h((\mathbf{X}, T)) - \epsilon}, 2^{h((\mathbf{X}, T)) + \epsilon} \right).$$

Krieger generator theorem If $H((\mathbf{X}, T)) < \infty$ then there exists a finite partition P such that $(P)_T = \mathcal{B}$. Thus every measure preserving transformation with finite entropy is isomorphic to a shift map on finitely many symbols (Krieger 1970).

Measure-theoretic entropy Entropy was introduced in physics by Rudolph Clausius in 1854. In 1948 Claude Shannon introduced the

concept to information theory. Consider a process that generates a string of data of length n . The entropy of the process is the smallest number h such that you can condense the data to a string of zeroes and one of length hn and with high probability you can reconstruct the original data from the string of zeroes and ones. Thus the entropy of a process is the average amount of information transmitted per symbol of the process.

Kolmogorov and Sinai introduced the concept of entropy to ergodic theory in the following way (Kolmogorov 1958, 1959). They defined the entropy of a partition Q is defined to be

$$H(Q) = - \sum_{i=1}^k \mu(Q_i) \log \mu(Q_i).$$

The measure-theoretic entropy of a dynamical system (X, T) with respect to a partition $Q : X \rightarrow \{1, \dots, k\}$ is then defined as

$$h(X, T, Q) = \lim_{N \rightarrow \infty} \frac{1}{N} H\left(\bigvee_{n=1}^N T^{-n} Q\right).$$

Finally, the measure-theoretic entropy of a dynamical system (X, T) is defined as

$$h((X, T)) = \sup_Q h(X, T, Q)$$

where the supremum is taken over all finite measurable partitions. A theorem of Sinai showed that if Q is a generator of (X, T) then $h(T) = h(T, Q)$ (Sinai 1959).

This shows that for every measure preserving function (X, T) there is an associated entropy $h(T) \in [0, \infty]$. It is easy to show from the definition that entropy is an isomorphism invariant.

We say that (Y, S) is a **factor** of (X, T) if there exists a map $\phi : X \rightarrow Y$ such that

1. ϕ is measure preserving and
2. $\phi(T(x)) = S(\phi(x))$ for almost every x .

Each factor (Y, S) can be associated with $\phi^{-1}(C)$, which is an invariant sub σ -algebra

of $\mathcal{B}$. We say that (Y, S) is trivial if Y consists of only one point. We say that a transformation (X, T) has **completely positive entropy** if every non-trivial factor of (X, T) has positive entropy.

Isomorphism of Bernoulli Shifts

Kolmogorov–Sinai

A long standing open question was for which $\mathbf{p}$ and $\mathbf{q}$ are Bernoulli $_p$ and Bernoulli $_q$ isomorphic. In particular are the Bernoulli 2 shift and the Bernoulli 3 shift isomorphic. Both of these transformations have completely positive entropy and all other isomorphism invariants which were known at the time are the same for the two transformations. The first application of the Kolmogorov–Sinai entropy was to show that the answer to this question is no.

Fix a probability vector $\mathbf{p}$. The transformation Bernoulli $_p$ has $Q_{\mathbf{p}} : x \rightarrow x_0$ as a generating partition. By Sinai's theorem

$$H(\text{Bernoulli}_p) = H(Q_{\mathbf{p}}) = \sum_{i=1}^n -p_i \log_2(p_i).$$

Thus the Bernoulli 2 shift (with entropy 1) is not isomorphic to the Bernoulli 3 shift (with entropy $\log_2(3)$).

Sinai also made significant progress toward showing that Bernoulli shifts with the same entropy are isomorphic by proving the following theorem.

Theorem 1 (Sinai 1962) *If (X, T) is a measure preserving system of entropy h and (Y, S) is a Bernoulli shift of entropy $h' \leq h$ then (Y, S) is a factor of (X, T) .*

This theorem implies that if $\mathbf{p}$ and $\mathbf{q}$ are probability vectors and $H(\mathbf{p}) = H(\mathbf{q})$ then Bernoulli $_p$ is a factor in Bernoulli $_q$ and Bernoulli $_q$ is a factor in Bernoulli $_p$. Thus we say that Bernoulli $_p$ and Bernoulli $_q$ are **weakly isomorphic**.

Explicit Isomorphisms

The other early progress on proving that Bernoulli shifts with the same entropy are isomorphic came

from Meshalkin. He considered pairs of probability vectors $\mathbf{p}$ and $\mathbf{q}$ with $H(\mathbf{p}) = H(\mathbf{q})$ and all of the p_i and q_i are related by some algebraic relations. For many such pairs he was able to prove that the two Bernoulli shifts are isomorphic. In particular he proved the following theorem.

Theorem 2 (Meshalkin 1959) Let $\mathbf{p} = (1/4, 1/4, 1/4, 1/4)$ and $\mathbf{q} = (1/2, 1/8, 1/8, 1/8, 1/8)$. Then Bernoulli_p and Bernoulli_q are isomorphic.

Ornstein

The central theorem in the study of isomorphisms of measure preserving transformations is Ornstein's isomorphism theorem.

Theorem 3 (Ornstein 1970c) If $\mathbf{p}$ and $\mathbf{q}$ are probability vectors and $H(\mathbf{p}) = H(\mathbf{q})$ then the Bernoulli shifts Bernoulli_p and Bernoulli_q are isomorphic.

To see how central this is to the field most of the rest of this article is a summary of:

1. The proof of Ornstein's theorem,
2. The consequences of Ornstein's theorem,
3. The generalizations of Ornstein's theorem, and
4. How the properties that Ornstein's theorem implies that Bernoulli shifts must have differ from the properties of every other class of transformations.

The key to Ornstein's proof was the introduction of the finitely determined property. To explain the finitely determined property we first define the Hamming distance of length n between sequences $x, y \in \mathbb{N}^{\mathbb{Z}}$ by

$$\bar{d}_n(x, y) = 1 - \frac{|\{k \in \{1, \dots, n\} : x_k = y_k\}|}{n}.$$

Let $(\mathbf{X}, T)$ and $(\mathbf{Y}, S)$ be a measure preserving transformation and let P and Q be finite partitions of X and Y respectively.

We say that $(\mathbf{X}, T)$ and P and $(\mathbf{Y}, S)$ and Q are within δ in n distributions if

$$\sum_{(a_1, \dots, a_n) \in \mathbb{Z}^n} |\mu(\{x : P(T^i(x)) = a_i \ \forall i = 1, \dots, n\}) - \nu(\{y : Q(T^i(y)) = a_i \ \forall i = 1, \dots, n\})| < \delta.$$

A process $(\mathbf{X}, T)$ and P are **finitely determined** if for every ϵ there exist n and δ such that if $(\mathbf{Y}, S)$ and Q are such that

1. $(\mathbf{X}, T)$ and P and $\mathbf{Y}$ and Q are within δ in n distributions and
 2. $|H(X, P) - H(Y, Q)| < \delta$
- then there exists a joining γ of X and Y such that for all m

$$\int_{x, y} \bar{d}_m(x, y) d\gamma(x, y) < \epsilon.$$

A transformation $(\mathbf{X}, T)$ is finitely determined if it is finitely determined for every finite partition P .

It is fairly straightforward to show that Bernoulli shifts are finitely determined. Ornstein used this fact along with the Rokhlin lemma and the Shannon–McMillan–Breiman theorem to prove a more robust version of Theorem 1.

To describe Ornstein's proof we use the a description due to Rothstein. We say that for a joining γ of (X, μ) and (Y, ν) that $P \subset_{\epsilon, \gamma} C$ if there exists a partition $\tilde{P}$ of C such that

$$\sum_i \gamma(P_i \Delta \tilde{P}_i) < \epsilon.$$

If $P \subset_{\epsilon, \gamma} C$ for all $\epsilon < 0$ then it is possible to show that there exists a partition $\tilde{P}$ of C such that

$$\sum_i \gamma(P_i \Delta \tilde{P}_i) = 0$$

and we write $P \subset_{\gamma} C$. If $P \subset_{\gamma} C$ then $(\mathbf{X}, T)$ is a factor of $(\mathbf{Y}, S)$ by the map ϕ that sends $y \rightarrow x$ where $P(T^i x) = \tilde{P}(S^i y)$ for all i .

In this language Ornstein proved that if $(\mathbf{X}, T)$ is finitely determined, P is a generating partition of X and $h((\mathbf{Y}, S)) \geq h((\mathbf{X}, T))$ then for every

$\epsilon > 0$ the set of joinings γ such that $P \subset_{\gamma, \epsilon} C$ is an open and dense set. Thus by the Baire category theorem there exists γ such that $P \subset_{\gamma} C$. This reproves Theorem 1.

Moreover if (X, T) and (Y, S) are finitely determined, $h((X, T)) = h((Y, S))$ and P and Q are generating partition of X and Y then by the Baire category theorem there exists γ such that $P \subset_{\gamma} C$ and $Q \subset_{\gamma} B$. Then the map ϕ that sends $y \rightarrow x$ where $P(T^i x) = \tilde{P}(S^i y)$ for all i is an isomorphism.

Properties of Bernoullis

Now we define the very weak Bernoulli property which is the most effective property for showing that a measure preserving transformation is isomorphic to a Bernoulli shift.

Given X and a partition P define the past of x by

$$P_{\text{past}}(x) = \{x' : T^i P(x') = T^i P(x) \ \forall i \in \mathbb{N}\}$$

and denote the measure μ conditioned on $P_{\text{past}}(x)$ by $\mu|_{P_{\text{past}}(x)}$. Define

$$\bar{d}_{n,x,\mu} = \inf_{\gamma} \int \bar{d}_n(x, x') d\gamma(x, x'),$$

where the inf is taken over all γ which are couplings of $\mu|_{P_{\text{past}}(x)}$ and μ . Also define

$$\bar{d}_{n,P_{\text{past}}(X,T)} = \int \bar{d}_{n,x,\mu} d\mu.$$

We say that (X, T) and P are very weak Bernoulli if for every $\epsilon > 0$ there exists n such that $\bar{d}_{n,P_{\text{past}}(X,T)} < \epsilon$. We say that (X, T) is very weakly Bernoulli if there exists a generating partition P such that (X, T) and P are very weak Bernoulli.

Ornstein and Weiss were able to show that the very weak Bernoulli property is both necessary and sufficient to be isomorphic to a Bernoulli shift.

Theorem 4 (Ornstein 1970b; Ornstein and Weiss 1974) *For transformations (X, T) the following conditions are equivalent:*

1. (X, T) is finitely determined,
2. (X, T) is very weak Bernoulli and
3. (X, T) is isomorphic to a Bernoulli shift.

Using the fact that a transformation is finitely determined or very weak Bernoulli is equivalent to it being isomorphic to a Bernoulli shift we can prove the following theorem.

Theorem 5 (Ornstein 1970a)

1. If (X, T^n) is isomorphic to a Bernoulli shift then (X, T) is isomorphic to a Bernoulli shift.
2. If (X, T) is a factor of a Bernoulli shift then (X, T) is isomorphic to a Bernoulli shift.
3. If (X, T) is isomorphic to a Bernoulli shift then there exists a measure preserving transformation (Y, S) such that (Y, S^n) is isomorphic to (X, T) .

Rudolph Structure Theorem

An important application of the very weak Bernoulli condition is the following theorem of Rudolph.

Theorem 6 (Rudolph 1983) *Let (X, T) be isomorphic to a Bernoulli shift, G be a compact Abelian group with Haar measure μ_G and $\sigma : X \rightarrow G$ be a measurable map. Then let $S : X \times G \rightarrow X \times G$ be defined by*

$$S(x, g) = (T(x), g + \sigma(x)).$$

Then $(X \times G, S, \mu \times \mu_G)$ is isomorphic to a Bernoulli shift.

Transformations Isomorphic to Bernoulli Shifts

One of the most important features of Ornstein's isomorphism theory is that it can be used to check whether specific transformations (or families of transformations) are isomorphic to Bernoulli shifts. The finitely determined property is the key to the proof of Ornstein's theorem and the

proof of many of the consequences listed in section “[Properties of Bernoullis](#)”. However if one wants to show a particular transformation is isomorphic to a Bernoulli shift then the very weak Bernoulli property is more useful. There have been many classes of transformations that have been proven to be isomorphic to a Bernoulli shift. Here we mention two.

The first class are the Markov chains. Friedman and Ornstein proved that if a Markov chain is mixing then it is isomorphic to a Bernoulli shift (Friedman and Ornstein 1970). The second are automorphisms of $[0, 1]^n$. Let M be any $n \times n$ matrix with integer coefficients and $|\det(M)| = 1$. If none of the eigenvalues $\{\lambda_i\}_{i=1}^n$ of M are roots of unity then Katznelson proved that T_M is isomorphic to the Bernoulli shift with entropy $\sum_{i=1}^n \max(0, \log(\lambda_i))$ (Katznelson 1971).

Transformations Not Isomorphic to Bernoulli Shifts

Recall that a measure preserving transformation (X, T) has **completely positive entropy** if for every nontrivial (Y, S) which is a factor of (X, T) we have that $H((Y, S)) > 0$. It is easy to check that Bernoulli shifts have completely positive entropy. It is natural to ask if the converse true? We shall see that the answer is an emphatic no. While the isomorphism class of Bernoulli shifts is given by just one number, the situation for transformations of completely positive entropy is infinitely more complicated.

Ornstein constructed the first example of a transformation with completely positive entropy which is not isomorphic to a Bernoulli shift (Ornstein 1973b). Ornstein and Shields built upon this construction to prove the following theorem.

Theorem 7 (Ornstein and Shields 1973) *For every $h > 0$ there is an uncountable family of completely positive entropy transformations which all have entropy h but no two distinct members of the family are isomorphic.*

Now that we see there are many isomorphic transformations that have completely positive

entropy it is natural to ask if (X, T) is not isomorphic to a Bernoulli shift then is there any reasonable condition we can put on (Y, S) that implies the two transformations are isomorphic. For example if (X, T^2) and (Y, S^2) are completely positive entropy transformations which are isomorphic does that necessarily imply that (X, T) and (Y, S) are isomorphic? The answer turns out to be no (Rudolph 1976). We could also ask if (X, T) and (Y, S) are completely positive entropy transformations which are weakly isomorphic does that imply that (X, T) and (Y, S) are isomorphic? Again the answer is no (Hoffman 1999a). The key insight to answering questions like this is due to Rudolph who showed that such questions about the isomorphism of transformations can be reduced to questions about conjugacy of permutations.

Rudolph’s Counterexample Machine

Given any transformation (X, T) and any permutation π in S_n (or $S_{\mathbb{N}}$) we can define the transformation $(X^n, T_\pi, \mu^n, \mathcal{B})$ by

$$T_\pi(x_1, x_2, \dots, x_n) = \left(T(x_{\pi(1)}), T(x_{\pi(2)}), \dots, T(x_{\pi(n)}) \right)$$

where μ^n is the direct product of n copies of μ . Rudolph introduced the concept of a transformation having **minimal self joinings**. If a transformation has minimal self joinings then for every π it is possible to list all of the measures on X^n which are invariant under T_π .

If there exists an isomorphism ϕ between T and (Y, S) then there is a corresponding measure on $X \times Y$ which is supported on points of the form $(x, \phi(x))$ and has marginals μ and ν . Thus if we know all of the measures on $X \times Y$ which are invariant under $T \times S$ then we know all of the isomorphisms between (X, T) and (Y, S) . Using this we get the following theorem.

Theorem 8 (Rudolph 1979) *There exists a non-trivial transformation with minimal self joinings. For any transformation (X, T) with minimal self joinings the corresponding transformation T_{π_1} is*

isomorphic to T_{π_2} if and only if the permutations π_1 and π_2 are conjugate.

There are two permutations on two elements, the flip $\pi_1 = (12)$ and the identity $\pi_2 = (1)(2)$. For both permutations, the square of the permutation is the identity. Thus there are two distinct permutations whose square is the same. Rudolph showed that this fact can be used to generate two transformations which are mixing that are not isomorphic but their squares are isomorphic. The following theorem gives more examples of the power of this technique.

Theorem 9 (Rudolph 1979)

1. *There exists measure preserving transformations (X, T) and (Y, S) which are weakly isomorphic but not isomorphic.*
2. *There exists measure preserving transformations (X, T) and (Y, S) which are not isomorphic but (X, T^k, μ) is isomorphic to (Y, S^k, ν) for every $k > 1$, and*
3. *There exists a mixing transformation with no non trivial factors.*

If (X, T) has minimal self joinings then it has zero entropy. However Hoffman constructed a transformation with completely positive entropy that shares many of the properties of transformations with minimal self joinings listed above.

Theorem 10 (Hoffman 1999a)

1. *There exist measure preserving transformations (X, T) and (Y, S) which both have completely positive entropy and are weakly isomorphic but not isomorphic.*
2. *There exist measure preserving transformations (X, T) and (Y, S) which both have completely positive entropy and are not isomorphic but (X, T^k, μ) is isomorphic to (Y, S^k, ν) for every $k > 1$.*

T, T Inverse

All of the transformations that have completely positive entropy but are not isomorphic to a Bernoulli shift described above are constructed by a process called cutting and stacking. These

transformations have little inherent interest outside of their ergodic theory properties. This led many people to search for a “natural” example of such a transformation. The most natural examples are the $[T, T^{-1}]$ transformation and many other transformations derived from it. It is easy to show that the $[T, T^{-1}]$ transformation has completely positive entropy (Meilijson 1974). Kalikow proved that for many T the corresponding $[T, T^{-1}]$ transformation is not isomorphic to a Bernoulli shift.

Theorem 11 (Kalikow 1982) *If $h(T) > 0$ then the $[T, T^{-1}]$ transformation is not isomorphic to a Bernoulli shift.*

The basic idea of Kalikow’s proof has been used by many others. Katok and Rudolph used the proof to construct smooth measure preserving transformations on infinite differentiable manifolds which have completely positive entropy but are not isomorphic to Bernoulli shifts (Katok 1980; Rudolph 1988). Den Hollander and Steif did a thorough study of the ergodic theory properties of $[T, T^{-1}]$ transformations where T is simple random walk on a wide family of graphs (Den Hollander and Steif 1997).

Classifying the Invariant Measures of Algebraic Actions

This problem of classifying all of the invariant measures like Rudolph did with his transformation with minimal self joinings comes up in a number of other settings. Ratner characterized the invariant measures for the horocycle flow and thus characterized the possible isomorphisms between a large family of transformations generated by the horocycle flow (Ratner 1978, 1982, 1983). This work has powerful applications to number theory.

There has also been much interest in classifying the measures on $[0, 1)$ that are invariant under both times 2 and times 3. Furstenberg proved that the only closed infinite set on the circle which is invariant under times 2 and times 3 is $[0, 1)$ itself and made the following measure theoretic conjecture.

Conjecture 1 (Furstenberg 1967) *The only non-atomic measure on $[0, 1)$ which is invariant under times 2 and times 3 is Lebesgue measure.*

Rudolph improved on the work of Lyons (1988) to provide the following partial answer to this conjecture.

Theorem 12 (Rudolph 1990) *The only measure on $[0, 1)$ which is invariant under multiplication by 2 and by 3 and has positive entropy under multiplication by 2 is Lebesgue measure.*

Johnson then proved that for all relatively prime p and q that p and q can be substituted for 2 and 3 in the theorem above.

This problem can be generalized to higher dimensions by studying the actions of commuting integer matrices of determinant greater than one on tori. Katok and Spatzier (1996) and Einsiedler and Lindenstrauss (2003) obtained results similar to Rudolph's for actions of commuting matrices.

Finitary Isomorphisms

By Ornstein's theorem we know that there exists an isomorphism between any two Bernoulli shifts (or mixing Markov shifts) of the same entropy. There has been much interest in studying how "nice" the isomorphism can be. By this we mean can ϕ be chosen so that the map $x \rightarrow (\phi(x))_0$ is continuous and if so what is its best possible modulus of continuity?

We say that a map ϕ from $\mathbb{N}^{\mathbb{Z}}$ to $\mathbb{N}^{\mathbb{Z}}$ is **finitary** if in order to determine $(\phi(x))_0$ we only need to know finitely many coordinates of x . More precisely if for almost every x there exists $m(x)$ such that

$$\begin{aligned} & \mu \left(\left\{ x' : x'_i = x_i \text{ for all } |i| \leq m(x) \text{ and } (\phi(x'))_0 \neq \phi(x)_0 \right\} \right) \\ &= 0. \end{aligned}$$

We say that $m(x)$ has **finitary moment** if $\int m(x)^t d\mu < \infty$ and that ϕ has **finite expected coding length** if the first moment of $m(x)$ is finite.

Keane and Smorodinsky proved the following strengthening of Ornstein's isomorphism theorem.

Theorem 13 (Keane and Smorodinsky 1979) *If $\mathbf{p}$ and $\mathbf{q}$ are probability vectors with $H(\mathbf{p}) = H(\mathbf{q})$ then the Bernoulli shifts $\text{Bernoulli}_{\mathbf{p}}$ and $\text{Bernoulli}_{\mathbf{q}}$ are isomorphic and there exists an isomorphism ϕ such that ϕ and ϕ^{-1} are both finitary.*

The nicest that we could hope ϕ to be is if both ϕ and ϕ^{-1} are finitary and have finite expected coding length. Schmidt proved that this happens only in the trivial case that $\mathbf{p}$ and $\mathbf{q}$ are rearrangements of each other.

Theorem 14 (Schmidt 1984) *If $\mathbf{p}$ and $\mathbf{q}$ are probability vectors and the Bernoulli shifts $\text{Bernoulli}_{\mathbf{p}}$ and $\text{Bernoulli}_{\mathbf{q}}$ are isomorphic and there exists an isomorphism ϕ such that ϕ and ϕ^{-1} are both finitary and have finite expected coding time then $\mathbf{p}$ is a rearrangement of $\mathbf{q}$.*

The best known result about this problem is the following theorem of Harvey and Peres.

Theorem 15 (Harvey and Peres) *If $\mathbf{p}$ and $\mathbf{q}$ are probability vectors with $H(\mathbf{p}) = H(\mathbf{q})$ then $\sum_i (p_i)^2 \log(p_i) = \sum_i (q_i)^2 \log(q_i)$ if and only if the Bernoulli shifts $\text{Bernoulli}_{\mathbf{p}}$ and $\text{Bernoulli}_{\mathbf{q}}$ are isomorphic and there exists an isomorphism ϕ such that ϕ and ϕ^{-1} are both finitary and have finite one half moment.*

Flows

A **flow** is a measure preserving action of the group $\mathbb{R}$ on a measure space $(\mathbf{X}, T)$. A **cross section** is any measurable set in $C \subset X$ such that for almost every x

$$0 < \inf \{t : T_t(x) \in C\} < \infty.$$

For any flow $(\mathbf{X}, T)$ and $\{T_t\}_{t \in \mathbb{R}}$ and any cross section C we define the **return time map for C** $R : C \rightarrow \mathbb{R}$ as follows. For any $x \in C$ define

$$t(x) = \inf\{t : T_t(x) \in C\}$$

then set $R(x) = T_{t(x)}(x)$. There is a standard method to project the probability measure μ on X to an invariant probability measure μ_C on C as well as the σ -algebra $\mathcal{B}$ on X to a σ -algebra $\mathcal{B}_C$ on C such that $(C, \mu_C, \mathcal{B}_C)$ and R is a measure preserving transformation.

First we show that there is a natural analog of Bernoulli shifts for flows.

Theorem 16 (Ornstein 1970b) *There exists a flow $(\mathbf{X}, T)$ and $\{T_t\}_{t \in \mathbb{R}}$ such that for every $t > 0$ the map $(\mathbf{X}, T)$ and T_t is isomorphic to a Bernoulli shift. Moreover for any $h \in (0, \infty]$ there exists $(\mathbf{X}, T)$ and $\{T_t\}_{t \in \mathbb{R}}$ such that $h(T_1) = h$.*

We say that such a flow $(X, \{f_t\}_{t \in \mathbb{R}}, \mu, \mathcal{B})$ is a **Bernoulli flow**.

This next version of Ornstein's isomorphism theorem shows that up to isomorphism and a change in time (considering the flow $\mathbf{X}$ and $\{T_{ct}\}$ instead of $\mathbf{X}$ and $\{T_t\}$) there are only two Bernoulli flows, one with positive but finite entropy and one with infinite entropy.

Theorem 17 (Ornstein 1970b) *If $(\mathbf{X}, T)$ and $\{T_t\}_{t \in \mathbb{R}}$ and $(\mathbf{Y}, S)\{S_t\}_{t \in \mathbb{R}}$ are Bernoulli flows and $h(T_1) = h(S_1)$ then they are isomorphic.*

As in the case of actions of $\mathbb{Z}$ there are many natural examples of flows that are isomorphic to the Bernoulli flow. The first is for geodesic flows. In the 1930s Hopf proved that geodesic flows on compact surfaces of constant negative curvature are ergodic (Hopf 1971). Ornstein and Weiss extended Hopf's proof to show that the geodesic flow is also Bernoulli (Ornstein and Weiss 1973).

The second class of flows comes from billiards on a square table with one circular bumper. The state space X consists of all positions and velocities for a fixed speed. The flow T_t is frictionless movement for time t with elastic collisions. This flow is also isomorphic to the Bernoulli flow (Gallavotti and Ornstein 1974).

Other Equivalence Relations

In this section we will discuss a number of equivalence relations between transformations that are weaker than isomorphism. All of these equivalence relations have a theory that is parallel to Ornstein's theory.

Kakutani Equivalence

We say that two transformations $(\mathbf{X}, T)$ and $(\mathbf{Y}, S)$ are **Kakutani equivalent** if there exist subsets $A \subset X$ and $B \subset Y$ such that (T_A, A, μ_A) and (S_B, B, ν_B) are isomorphic. This is equivalent to the existence of a flow $(\mathbf{X}, T)$ and $\{T_t\}_{t \in \mathbb{R}}$ with cross sections C and C' such that the return time maps of C and C' are isomorphic to $(\mathbf{X}, T)$ and $(\mathbf{Y}, S)$ respectively.

Using the properties of entropy of the induced map we have that if $(\mathbf{X}, T)$ and $(\mathbf{Y}, S)$ are Kakutani equivalent then either $h((\mathbf{X}, T)) = h((\mathbf{Y}, S)) = 0$, $0 < h((\mathbf{X}, T)), h((\mathbf{Y}, S)) < \infty$ or $h((\mathbf{X}, T)) = h((\mathbf{Y}, S)) = \infty$.

In general the answer to the question of which pairs of measure preserving transformations are isomorphic is quite poorly understood. But if one of the transformations is a Bernoulli shift then Ornstein's theory gives a fairly complete answer to the question. A similar situation exists for Kakutani equivalence. In general the answer to the question of which pairs of measure preserving transformation are Kakutani equivalent is also quite poorly understood. But the more specialized question of which transformations are isomorphic to a Bernoulli shift has a more satisfactory answer.

Feldman constructed a transformation $(\mathbf{X}, T)$ which has completely positive entropy but $(\mathbf{X}, T)$ is not Kakutani equivalent to a Bernoulli shift. Ornstein, Rudolph and Weiss extended Feldman's work to construct a complete theory of the transformations that are Kakutani equivalent to a Bernoulli shift (Ornstein et al. 1982) for positive entropy transformations and a theory of the transformations that are Kakutani equivalent to an irrational rotation (Ornstein et al. 1982) for zero entropy transformations. (The zero entropy version of this theorem had been developed independently (and earlier) by Katok (1975).)

They defined two class of transformations called **loosely** Bernoulli and **finitely fixed**. The definitions of these properties are the same as the definitions of very weak Bernoulli and finitely determined except that the $\bar{d}$ metric is replaced by the $\bar{f}$ metric. For $x, y \in \mathbb{N}^{\mathbb{Z}}$ we define

$$\bar{f}_n(x, y) = 1 - \frac{k}{n}$$

where k is the largest number such that sequences $1 \leq i_1 < i_2 < \dots < i_k \leq n$ and $1 \leq j_1 < j_2 < \dots < j_k \leq n$ such that $x_{i_j} = y_{j_i}$ for all j , $1 \leq j \leq k$. (In computer science this metric is commonly referred to as the edit distance.) Note that $\bar{d}_n(x, y) \geq \bar{f}_n(x, y)$.

They proved the following analog of Theorem 5.

Theorem 18 *For transformations $(\mathbf{X}, T)$ with $h(\mathbf{X}, T) > 0$ the following conditions are equivalent:*

1. $(\mathbf{X}, T)$ is finitely fixed,
2. $(\mathbf{X}, T)$ is loosely Bernoulli,
3. $(\mathbf{X}, T)$ is Kakutani equivalent to a Bernoulli shift and
4. There exists a Bernoulli flow $\mathbf{Y}$ and $\{F_t\}_{t \in \mathbb{R}}$ and a cross section C such that the return time map for C is isomorphic to $(\mathbf{X}, T)$.

Restricted Orbit Equivalence

Using the $\bar{d}$ metric we got a theory of which transformations are isomorphic to a Bernoulli shift. Using the $\bar{f}$ metric we got a strikingly similar theory of which transformations are Kakutani equivalent to a Bernoulli shift. Rudolph showed that it is possible to replace the $\bar{d}$ metric (or the $\bar{f}$ metric) with a wide number of other metrics and produce parallel theories for other equivalence relations. For instance, for each of these theories we get a version of Theorem 5. This collection of theories is called restricted orbit equivalence (Rudolph 1985).

Non-invertible Transformations

The question of which noninvertible measure preserving transformations are isomorphic turns out

to be quite different from the same question for invertible transformations. In one sense it is easier because of an additional isomorphism invariant.

For any measure preserving transformation $(\mathbf{X}, T)$ the probability measure $\mu|_{T^{-1}(x)}$ on $T^{-1}(x)$ is defined for almost every $x \in X$. (If $(\mathbf{X}, T)$ is invertible then this measure is trivial as $|\{T^{-1}(x)\}| = 1$ and $\mu|_{T^{-1}(x)}(T^{-1}(x)) = 1$ for almost every x .) It is easy to check that if ϕ is an isomorphism from $(\mathbf{X}, T)$ to $(\mathbf{Y}, S)$ then for almost every x and $x' \in T^{-1}(x)$ we have

$$\mu|_{T^{-1}(x)}(x') = \nu|_{S^{-1}(\phi(x))}(\phi(x')).$$

From this we can easily see that if $\mathbf{p} = \{p_i\}_{i=1}^n$ and $\mathbf{q} = \{q_i\}_{i=1}^m$ are probability vectors then the corresponding one sided Bernoulli shifts are isomorphic only if $m = n$ and there is a permutation $\pi \in S_n$ such that $p_{\pi(i)} = q_i$ for all i . (In this case we say $\mathbf{p}$ is a **rearrangement** of $\mathbf{q}$.) If $\mathbf{p}$ is a rearrangement of $\mathbf{q}$ then it is easy to construct an isomorphism between the corresponding Bernoulli shifts. Thus the analog of Ornstein's theorem for Bernoulli endomorphisms is trivial. However we will see that there still is an analogous theory classifying the class of endomorphism that are isomorphic to Bernoulli endomorphisms.

We say that an endomorphism is **uniformly d to 1** if for almost every x we have that $|\{T^{-1}(x)\}| = d$ and $\mu|_{T^{-1}(x)}(y) = 1/d$ for all $y \in T^{-1}(x)$. Hoffman and Rudolph defined two classes of noninvertible transformations called **tree very weak Bernoulli** and **tree finitely determined** and proved the following theorem.

Theorem 19 *The following three conditions are equivalent for uniformly d to 1 endomorphisms.*

1. $(\mathbf{X}, T)$ is tree very weak Bernoulli,
2. $(\mathbf{X}, T)$ is tree finitely determined, and
3. $(\mathbf{X}, T)$ is isomorphic to the one sided Bernoulli d shift.

Jong extended this theorem to say that if there exists a probability vector $\mathbf{p}$ such that for almost every x the distribution of $\mu|_{T^{-1}(x)}$ is the same as the distribution of $\mathbf{p}$ then $(\mathbf{X}, T)$ is isomorphic to the one sided Bernoulli d shift if and only if it is

tree finitely determined and if and only if it is tree very weak Bernoulli (Jong 2003).

Markov Shifts

We saw that mixing Markov chains are isomorphic if they have the same entropy. As we have seen there are additional isomorphism invariants for noninvertible transformations. Ashley, Marcus and Tuncel managed to classify all one sided mixing Markov chains up to isomorphism (Ashley et al. 1997).

Rational Maps

Rational maps are the main object of study in complex dynamics. For every rational function $f(z) = p(z)/q(z)$ there is a nonempty compact set J_f which is called the **Julia set**. Roughly speaking this is the set of points for which every neighborhood acts “chaotically” under repeated iterations of f .

In order to consider rational maps as measure preserving transformations we need to specify an invariant measure. The following theorem of Gromov shows that for every rational map there is one canonical measure to consider.

Theorem 20 (Gromov 2003) *For every f rational function of degree d there exists a unique invariant measure μ_f of maximal entropy. We have that $h(\mu_f) = \log_2 d$ and $\mu_f(J_f) = 1$.*

The properties of this measure were studied by Freire, Lopes and Mañé (1983). Mañé that analysis to prove the following theorem.

Theorem 21 (Mañé 1985) *For every rational function f of degree d there exists n such that $(\mathbb{C}, f^n, \mu_f)$ (where $f^n(z) = f(f(f \dots (z)))$) is composition) is isomorphic to the one sided Bernoulli d^n shift.*

Heicklen and Hoffman used the tree very weak Bernoulli condition to show that we can always take n to be one.

Theorem 22 (Heicklen and Hoffman 2002) *For every rational function f of degree $d \geq 2$ the*

corresponding map $((\mathbb{C}, \mu_f, \mathcal{B}), T_f)$ is isomorphic to the one sided Bernoulli d shift.

Differences with Ornstein’s Theory

Unlike Kakutani equivalence and restricted orbit equivalence which are very close parallels to Ornstein’s theory, the theory of which endomorphisms are isomorphic to a Bernoulli endomorphism contains some significant differences. One of the principal results of Ornstein’s isomorphism theory is that if $(\mathbf{X}, T)$ is an invertible transformation and $(\mathbf{X}, T^2)$ is isomorphic to a Bernoulli shift then $(\mathbf{X}, T)$ is also isomorphic to a Bernoulli shift. There is no corresponding result for noninvertible transformations.

Theorem 23 (Hoffman 2004) *There is a uniformly two to one endomorphism $(\mathbf{X}, T)$ which is not isomorphic to the one sided Bernoulli 2 shift but $(\mathbf{X}, T^2)$ is isomorphic to the one sided Bernoulli 4 shift.*

Factors of a Transformation

In this section we study the relationship between a transformation and its factors. There is a natural way to associate a factor of $(\mathbf{X}, T)$ with a sub σ -algebra of $\mathcal{B}$. Let $(\mathbf{Y}, S)$ be a factor of $(\mathbf{X}, T)$ with factor map $\phi : X \rightarrow Y$. Then the σ -algebra associated with $(\mathbf{Y}, S)$ is $\mathcal{B}_Y = \phi^{-1}(\mathcal{C})$. Thus the study of factors of a transformation is the study of its sub σ -algebras.

Almost every property that we have discussed above has an analog in the study of factors of a transformation. We give three such examples. We say that two factors $\mathcal{C}$ and $\mathcal{D}$ of $(\mathbf{X}, T)$ are **relatively isomorphic** if there exists an isomorphism $\psi : X \rightarrow X$ of $(\mathbf{X}, T)$ with itself such that $\psi(\mathcal{C}) = \mathcal{D}$. We say that $(\mathbf{X}, T)$ has **relatively completely positive entropy with respect to $\mathcal{C}$** if every factor $\mathcal{D}$ which contains $\mathcal{C}$ has $h(\mathcal{D}) > h(\mathcal{C})$. We say that $\mathcal{C}$ is **relatively Bernoulli** if there exists a second factor $\mathcal{D}$ which is independent of $\mathcal{C}$ and $\mathcal{B} = \mathcal{C} \vee \mathcal{D}$.

Thouvenot defined properties of factors called **relatively very weak Bernoulli** and **relatively finitely determined**. Then he proved an analog

of Theorem 3. This says that a factor being relatively Bernoulli is equivalent to it being relatively finitely determined (and also equivalent to it being relatively very weak Bernoulli).

The **Pinsker algebra** is the maximal σ -algebra $\mathcal{P}$ such that $h(\mathcal{P}) = 0$. The Pinsker conjecture was that for every measure preserving transformation (X, T) there exists a factor $\mathcal{C}$ such that

1. $\mathcal{C}$ is independent of the Pinsker algebra $\mathcal{P}$
2. $\mathcal{B} = \mathcal{C} \vee \mathcal{P}$ and
3. $(X, \mathcal{C})$ has completely positive entropy.

Ornstein found a counterexample to the Pinsker conjecture (Ornstein 1973c). After Thouvenot developed the relative isomorphism theory he came up with the following question which is referred to as the weak Pinsker conjecture.

Conjecture 2 *For every measure preserving transformation (X, T) and every $\epsilon > 0$ there exist invariant σ -algebras $\mathcal{C}, \mathcal{D} \subset \mathcal{B}$ such that*

1. $\mathcal{C}$ is independent of $\mathcal{D}$
2. $\mathcal{B} = \mathcal{C} \vee \mathcal{D}$
3. $(X, T, \mu, \mathcal{D})$ is isomorphic to a Bernoulli shift
4. $h((X, T, \mu, \mathcal{C})) < \epsilon$.

There is a wide class of transformations which have been proven to satisfy the weak Pinsker conjecture. This class includes almost all measure preserving transformations which have been extensively studied.

Actions of Amenable Groups

All of the discussion above has been about the action of a single invertible measure preserving transformation (actions of $\mathbb{N}$ and $\mathbb{Z}$) or flows (actions of $\mathbb{R}$). We now consider more general group actions. If we have two actions S and T on a measure space (X, μ) which commute ($S(T(x)) = T(S(x))$ for almost every x) then there is an action of $\mathbb{Z}^2$ on (X, μ) given by $f_{(n,m)}(x) = S^n(T^m(x))$. A natural question to ask is do there exist a version of entropy theory and Ornstein's isomorphism theory for actions of two commuting automorphisms. More generally for each of the

results discussed above we can ask what is the largest class of groups such that an analogous result is true. It turns out that for most of the properties described above the right class of groups is discrete amenable groups.

A **Følner sequence** F_n in a group G is a sequence of subsets F_n of G such that for all $g \in G$ we have that $\lim_{n \rightarrow \infty} |g(F_n)| / |F_n| = 1$. A countable group is **amenable** if and only if it has a Følner sequence.

For nonamenable groups it is much more difficult to generalize Birkhoff's ergodic theorem (Nevo 1994; Nevo and Stein 1994). Lindenstrauss proved that for every discrete amenable group there is an analog of the ergodic theorem (Lindenstrauss 2001). For every amenable group G and every probability vector $\mathbf{p}$ we can define a Bernoulli action of G . There are also analogs of Rokhlin's lemma and the Shannon–McMillan–Breiman theorem for actions of all discrete amenable groups (Krieger 1970; Ornstein and Weiss 1983, 1987). Thus we have all of the ingredients to prove a version of Ornstein's isomorphism theorem.

Theorem 24 *If $\mathbf{p}$ and $\mathbf{q}$ are probability vectors and $H(\mathbf{p}) = H(\mathbf{q})$ then the Bernoulli actions of G corresponding to $\mathbf{p}$ and $\mathbf{q}$ are isomorphic.*

Also all of the aspects of Rudolph's theory of restricted orbit equivalence can be carried out for actions of amenable groups (Kammeyer and Rudolph 2002).

Differences Between Actions of $\mathbb{Z}$ and Actions of Other Groups

Although generalizations of Ornstein theory and restricted orbit equivalence carry over well to the actions of discrete amenable groups there do turn out to be some significant differences between the possible actions of $\mathbb{Z}$ and those of other discrete amenable groups.

Many of these have to do with the generalization of Markov shifts. For actions of $\mathbb{Z}^2$ these are called Markov random fields. By the result of Friedman and Ornstein if a Markov chain is mixing then it has completely positive entropy and it is isomorphic to a Bernoulli shift. Mixing Markov random fields can have very different properties. Ledrappier constructed a $\mathbb{Z}^2$ action

which is a Markov random field and is mixing but has zero entropy (Ledrappier 1978). Even more surprising even though it is mixing it is not mixing of all orders. The existence of a $\mathbb{Z}$ action which is mixing but not mixing of all orders is one of the longest standing open questions in ergodic theory (Halmos 1950).

Even if we try to strengthen the hypothesis of Friedman and Ornstein's theorem to assume that the Markov random field has completely positive entropy we will not succeed as there exists a Markov random field which has completely positive entropy but is not isomorphic to a Bernoulli shift (Hoffman 1999b).

Future Directions

In the future we can expect to see progress of isomorphism theory in a variety of different directions. One possible direction for future research is better understand the properties of finitary isomorphisms between various transformations and Bernoulli shifts described in section “[Finitary Isomorphisms](#)”. Another possible direction would be to find a theory of equivalence relations for Bernoulli endomorphisms analogous to the one for invertible Bernoulli transformations described in section “[Other Equivalence Relations](#)”.

As the subject matures the focus of research in isomorphism theory will likely shift to connections to other fields. Already there are deep connections between isomorphism theory and both number theory and statistical physics. Finally one hopes to see progress made on the two dominant outstanding conjectures in the field: Thouvenot weak Pinsker conjecture (Conjecture 2) and Furstenberg's conjecture (Conjecture 1) about measures on the circle invariant under both the times 2 and times 3 maps. Progress on either of these conjectures would invariably lead the field in exciting new directions.

Bibliography

- Ashley J, Marcus B, Tuncel S (1997) The classification of one-sided Markov chains. *Ergod Theory Dynam Syst* 17(2):269–295
- Birkhoff GD (1931) Proof of the ergodic theorem. *Proc Natl Acad Sci U S A* 17:656–660
- Breiman L (1957) The individual ergodic theorem of information theory. *Ann Math Stat* 28:809–811
- Den Hollander F, Steif J (1997) Mixing E properties of the generalized T, T^{-1} -process. *J Anal Math* 72:165–202
- Einsiedler M, Lindenstrauss E (2003) Rigidity properties of $\mathbb{Z}^d$ -actions on tori and solenoids. *Electron Res Announc Am Math Soc* 9:99–110
- Einsiedler M, Katok A, Lindenstrauss E (2006) Invariant measures and the set of exceptions to Littlewood's conjecture. *Ann Math* (2) 164(2):513–560
- Feldman J (1976) New K -automorphisms and a problem of Kakutani. *Isr J Math* 24(1):16–38
- Freire A, Lopes A, Mañé R (1983) An invariant measure for rational maps. *Bol Soc Brasil Mat* 14(1):45–62
- Friedman NA, Ornstein DS (1970) On isomorphism of weak Bernoulli transformations. *Adv Math* 5:365–394
- Furstenberg H (1967) Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math Syst Theory* 1:1–49
- Gallavotti G, Ornstein DS (1974) Billiards and Bernoulli schemes. *Commun Math Phys* 38:83–101
- Gromov M (2003) On the entropy of holomorphic maps. *Enseign Math* (2) 49(3–4):217–235
- Halmos PR (1950) *Measure theory*. D Van Nostrand, New York
- Halmos PR (1960) *Lectures on ergodic theory*. Chelsea Publishing, New York
- Harvey N, Peres Y. An invariant of finitary codes with finite expected square root coding length. *Ergod Theory Dynam Syst*, to appear
- Heicklen D (1998) Bernoullis are standard when entropy is not an obstruction. *Isr J Math* 107:141–155
- Heicklen D, Hoffman C (2002) Rational maps are d -adic Bernoulli. *Ann Math* (2) 156(1):103–114
- Hoffman C (1999a) A K counterexample machine. *Trans Am Math Soc* 351(10):4263–4280
- Hoffman C (1999b) A Markov random field which is K but not Bernoulli. *Isr J Math* 112:249–269
- Hoffman C (2003) The scenery factor of the $[T, T^{-1}]$ transformation is not loosely Bernoulli. *Proc Am Math Soc* 131(12):3731–3735
- Hoffman C (2004) An endomorphism whose square is Bernoulli. *Ergod Theory Dynam Syst* 24(2):477–494
- Hoffman C, Rudolph D (2002) Uniform endomorphisms which are isomorphic to a Bernoulli shift. *Ann Math* (2) 156(1):79–101
- Hopf E (1971) Ergodic theory and the geodesic flow on surfaces of constant negative curvature. *Bull Am Math Soc* 77:863–877
- Jong P (2003) On the isomorphism problem of p -endomorphisms. PhD thesis, University of Toronto
- Kalikow SA (1982) T, T^{-1} transformation is not loosely Bernoulli. *Ann Math* (2) 115(2):393–409
- Kammeyer JW, Rudolph DJ (2002) *Restricted orbit equivalence for actions of discrete amenable groups*. Cambridge tracts in mathematics, vol 146. Cambridge University Press, Cambridge
- Katok AB (1975) Time change, monotone equivalence, and standard dynamical systems. *Dokl Akad Nauk SSSR* 223(4):789–792. in Russian

- Katok A (1980) Smooth non-Bernoulli K -automorphisms. *Invent Math* 61(3):291–299
- Katok A, Spatzier RJ (1996) Invariant measures for higher-rank hyperbolic abelian actions. *Ergod Theory Dynam Syst* 16(4):751–778
- Katznelson Y (1971) Ergodic automorphisms of T^n are Bernoulli shifts. *Isr J Math* 10:186–195
- Keane M, Smorodinsky M (1979) Bernoulli schemes of the same entropy are finitarily isomorphic. *Ann Math* (2) 109(2):397–406
- Kolmogorov AN (1958) A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces. *Dokl Akad Nauk SSSR (NS)* 119: 861–864. in Russian
- Kolmogorov AN (1959) Entropy per unit time as a metric invariant of automorphisms. *Dokl Akad Nauk SSSR* 124:754–755. in Russian
- Krieger W (1970) On entropy and generators of measure-preserving transformations. *Trans Am Math Soc* 149: 453–464
- Ledrappier F (1978) Un champ markovien peut être d'entropie nulle et mélangeant. *CR Acad Sci Paris Sér A–B* 287(7):A561–A563. in French
- Lindenstrauss E (2001) Pointwise theorems for amenable groups. *Invent Math* 146(2):259–295
- Lyons R (1988) On measures simultaneously 2- and 3-invariant. *Isr J Math* 61(2):219–224
- Mañé R (1983) On the uniqueness of the maximizing measure for rational maps. *Bol Soc Brasil Mat* 14(1):27–43
- Mañé R (1985) On the Bernoulli property for rational maps. *Ergod Theory Dynam Syst* 5(1):71–88
- Meilijson I (1974) Mixing properties of a class of skew-products. *Isr J Math* 19:266–270
- Meshalkin LD (1959) A case of isomorphism of Bernoulli schemes. *Dokl Akad Nauk SSSR* 128:41–44. in Russian
- Nevo A (1994) Pointwise ergodic theorems for radial averages on simple Lie groups. I. *Duke Math J* 76(1):113–140
- Nevo A, Stein EM (1994) A generalization of Birkhoff's pointwise ergodic theorem. *Acta Math* 173(1):135–154
- Ornstein D (1970a) Factors of Bernoulli shifts are Bernoulli shifts. *Adv Math* 5:349–364
- Ornstein D (1970b) Two Bernoulli shifts with infinite entropy are isomorphic. *Adv Math* 5:339–348
- Ornstein D (1970c) Bernoulli shifts with the same entropy are isomorphic. *Adv Math* 4:337–352
- Ornstein DS (1973a) A K automorphism with no square root and Pinsker's conjecture. *Adv Math* 10:89–102
- Ornstein DS (1973b) An example of a Kolmogorov automorphism that is not a Bernoulli shift. *Adv Math* 10: 49–62
- Ornstein DS (1973c) A mixing transformation for which Pinsker's conjecture fails. *Adv Math* 10:103–123
- Ornstein DS, Shields PC (1973) An uncountable family of K -automorphisms. *Adv Math* 10:63–88
- Ornstein DS, Weiss B (1973) Geodesic flows are Bernoullian. *Isr J Math* 14:184–198
- Ornstein DS, Weiss B (1974) Finitely determined implies very weak Bernoulli. *Isr J Math* 17:94–104
- Ornstein D, Weiss B (1983) The Shannon–McMillan–Breiman theorem for a class of amenable groups. *Isr J Math* 44(1):53–60
- Ornstein DS, Weiss B (1987) Entropy and isomorphism theorems for actions of amenable groups. *J Anal Math* 48:1–141
- Ornstein DS, Rudolph DJ, Weiss B (1982) Equivalence of measure preserving transformations. *Mem Am Math Soc* 37(262) American Mathematical Society
- Parry W (1969) Entropy and generators in ergodic theory. WA Benjamin, New York
- Parry W (1981) Topics in ergodic theory. Cambridge tracts in mathematics, vol 75. Cambridge University Press, Cambridge
- Petersen K (1989) Ergodic theory. Cambridge studies in advanced mathematics, vol 2. Cambridge University Press, Cambridge
- Pinsker MS (1960) Dynamical systems with completely positive or zero entropy. *Dokl Akad Nauk SSSR* 133: 1025–1026. in Russian
- Ratner M (1978) Horocycle flows are loosely Bernoulli. *Isr J Math* 31(2):122–132
- Ratner M (1982) Rigidity of horocycle flows. *Ann Math* (2) 115(3):597–614
- Ratner M (1983) Horocycle flows, joinings and rigidity of products. *Ann Math* (2) 118(2):277–313
- Ratner M (1991) On Raghunathan's measure conjecture. *Ann Math* (2) 134(3):545–607
- Rudolph DJ (1976) Two nonisomorphic K -automorphisms with isomorphic squares. *Isr J Math* 23(3–4):274–287
- Rudolph DJ (1979) An example of a measure preserving map with minimal self-joinings, and applications. *J Anal Math* 35:97–122
- Rudolph DJ (1983) An isomorphism theory for Bernoulli free Z -skew-compact group actions. *Adv Math* 47(3): 241–257
- Rudolph DJ (1985) Restricted orbit equivalence. *Mem Am Math Soc* 54(323) American Mathematical Society
- Rudolph DJ (1988) Asymptotically Brownian skew products give non-loosely Bernoulli K -automorphisms. *Invent Math* 91(1):105–128
- Rudolph DJ (1990) $\times 2$ and $\times 3$ invariant measures and entropy. *Ergod Theory Dynam Syst* 10(2):395–406
- Schmidt K (1984) Invariants for finitary isomorphisms with finite expected code lengths. *Invent Math* 76(1): 33–40
- Shields P (1973) The theory of Bernoulli shifts. Chicago lectures in mathematics. University of Chicago Press, Chicago
- Sinai J (1959) On the concept of entropy for a dynamic system. *Dokl Akad Nauk SSSR* 124:768–771. in Russian
- Sinai JG (1962) A weak isomorphism of transformations with invariant measure. *Dokl Akad Nauk SSSR* 147: 797–800. in Russian
- Thouvenot J-P (1975a) Quelques propriétés des systèmes dynamiques qui se décomposent en un produit de deux systèmes dont l'un est un schéma de Bernoulli. Conference on ergodic theory and topological dynamics, Kibbutz, Lavi, 1974. *Isr J Math* 21(2–3):177–207. in French
- Thouvenot J-P (1975b) Une classe de systèmes pour lesquels la conjecture de Pinsker est vraie. Conference on ergodic theory and topological dynamics, Kibbutz Lavi, 1974. *Isr J Math* 21(2–3):208–214. in French

Dynamical Systems of Probabilistic Origin: Gaussian and Poisson Systems

Élise Janvresse¹, Emmanuel Roy² and Thierry De La Rue³

¹Laboratoire Amiénois de Mathématique Fondamentale et Appliquée, CNRS-UMR 7352, Université de Picardie Jules Verne, Amiens, France

²Laboratoire Analyse, Géométrie et Applications, Université Paris 13 Institut Galilée, Villetaneuse, France

³Laboratoire de Mathématiques Raphaël Salem, CNRS – Université de Rouen Normandie, Saint Étienne du Rouvray, France

Article Outline

Glossary
Definition of the Subject
Introduction
From Probabilistic Objects to Dynamical Systems
Spectral Theory
Basic Ergodic Properties
Joinings, Factors, and Centralizer
GAGs and PAPs
Future Directions
Bibliography

Glossary

Centralizer The *centralizer* of an invertible measure-preserving transformation T is the set $C(T)$ of all invertible measure-preserving transformations on the same measure space which commute with T .

(Simple) Counting measure A *counting measure* on a measurable space $(X, \mathcal{A})$ is a measure of the form $\sum_{i \in I} \delta_{x_i}$ where $(x_i)_{i \in I}$ is a countable family of elements of X . The counting measure is said to be *simple* if $x_i \neq x_j$ whenever $i \neq j$.

Gaussian process, Gaussian space A *Gaussian process* is a family of real-valued random variables defined on a probability space $(\Omega, \mathbb{P})$, such that any linear combination of finitely many of these random variables is either 0 or normally distributed.

A real linear subspace of $L^2(\mathbb{P})$ is a *Gaussian space* if any nonzero random variable it contains is normally distributed. The closure of the linear real subspace spanned by a Gaussian process is a Gaussian space.

Infinite divisibility Let $(G, \mathcal{G}, +)$ be a measurable Abelian semigroup, i.e., the addition $(G \times G, \mathcal{G} \otimes \mathcal{G}) \mapsto (G, \mathcal{G})$

$$(g_1, g_2) \mapsto g_1 + g_2$$

is commutative and measurable. The convolution $\nu * \rho$ of probability measures ν and ρ on $(G, \mathcal{G})$ is well defined as the image of $\nu \otimes \rho$ by the addition.

A probability measure ν on $(G, \mathcal{G})$ is *infinitely divisible* if for any $k \geq 1$, there exists a probability measure ν_k on $(G, \mathcal{G})$ such that $\nu = (\nu_k)^{*k}$.

Kronecker subset of the unidimensional torus A subset K of the unidimensional torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is a *Kronecker set* if any continuous function $f: K \rightarrow S^1$ is a uniform limit of characters: there exists a sequence $(k_n) \subset \mathbb{Z}$ such that

$$\max_{t \in K} |f(t) - e^{i2\pi k_n t}| \xrightarrow{n \rightarrow \infty} 0.$$

Any finite set of rationally independent elements of $\mathbb{T}$ is a Kronecker set, but there exist also perfect Kronecker subsets of $\mathbb{T}$ (see for example (Cornfeld et al. 1982), Appendix 4)).

Point process A *point process* N on $(X, \mathcal{A})$ is a random variable taking values in the space X^* of counting measures on $(X, \mathcal{A})$. It is said to be *simple* if N is almost surely a simple counting measure.

The measure $A \mapsto \mathbb{E}[N(A)]$ on $(X, \mathcal{A})$ is called the *intensity* of N .

A point process of intensity μ is said to have *moment of order* $k \geq 1$ if, for all $A \in \mathcal{A}$

with $0 < \mu(A) < \infty$,

$$\mathbb{E} [N(A)^k] < \infty.$$

Poisson point process Let $(X, \mathcal{A}, \mu)$ be a sigma-finite measure space. A *Poisson point process of intensity μ on X* is a point process N on X , such that

- For any set $A \in \mathcal{A}$ with $0 < \mu(A) < \infty$, $N(A)$ is a Poisson random variable of parameter $\mu(A)$.
- For any $k \geq 1$, for any collection $(A_1, \dots, A_k)$ in $\mathcal{A}$, the random variables $N(A_i)$, $1 \leq i \leq k$, are independent.

A Poisson process is simple if and only if its intensity is a continuous measure.

Self-joining and associated Markov operator A self-joining of the probability-preserving dynamical system $(X, \mathcal{A}, \mu, T)$ is a probability measure ν on the Cartesian square $(X \times X, \mathcal{A} \otimes \mathcal{A})$ such that

- Both marginals of ν are equal to μ .
- ν is $T \times T$ -invariant. When ν is such a self-joining, it gives rise to a bigger probability-preserving dynamical system $(X \times X, \mathcal{A} \otimes \mathcal{A}, \nu, T \times T)$ in which we see two copies of the original system as factors (via the projections on the two coordinates).

To a self-joining ν there corresponds a unique Markov operator Φ_ν (i.e., Φ_ν is a positive operator on $L^2(\mu)$ with $\Phi_\nu 1 = 1$) that commutes with the Koopman operator U_T (i.e., the unitary operator defined on $L^2(\mu)$ by $U_T h := h \circ T$), and which is characterized by the following relation: for any A and B in $\mathcal{A}$:

$$\nu(A \times B) = \langle 1_A, \Phi_\nu 1_B \rangle_{L^2(\mu)}.$$

Stationary process A *stationary process* is a sequence $(\xi_n)_{n \in \mathbb{Z}}$ of random variables taking values in a set V , such that for each $k \in \mathbb{Z}$ and any $i_1 < \dots < i_d \in \mathbb{Z}$, the distribution of $(\xi_{i_1+k}, \dots, \xi_{i_d+k})$ is the same as the distribution of $(\xi_{i_1}, \dots, \xi_{i_d})$. In other words, the process is stationary if its distribution is shift-invariant on $V^{\mathbb{Z}}$.

Spectral measure Let U be a unitary operator on a Hilbert space H . For any $h \in H$, there exists a finite positive measure σ_h on the

unidimensional torus, called the *spectral measure of h* , satisfying for all $k \in \mathbb{Z}$

$$\widehat{\sigma}_h(k) := \int_{\mathbb{T}} e^{-i2\pi kt} d\sigma_h(t) = \langle h, U^k h \rangle. \quad (1)$$

When T is an invertible measure-preserving transformation on a measure space $(X, \mathcal{A}, \mu)$, this applies in particular to the associated Koopman operator U_T . In the case of a square integrable stationary process $(\xi_n)_{n \in \mathbb{Z}}$ (see definition below), its distribution being shift-invariant, the covariances $\mathbb{E}[\xi_0 \overline{\xi_k}]$ of the process can be interpreted as scalar products $\langle h, U^k h \rangle$, where h is the projection on the 0-coordinate of the process, and U is the Koopman operator associated to the shift. Thus the spectral measure σ of the process satisfies, for all $k \in \mathbb{Z}$,

$$\widehat{\sigma}(k) = \mathbb{E}[\xi_0 \overline{\xi_k}]. \quad (2)$$

In the case of a real-valued stationary process, the covariances are real and this measure σ must be symmetric (invariant by $t \rightarrow -t$).

Definition of the Subject

Measure-theoretic dynamical systems are systems of the form $(X, \mathcal{A}, \mu, T)$, where $(X, \mathcal{A})$ is a standard Borel space, μ is a sigma-finite measure on $(X, \mathcal{A})$, and $T: X \rightarrow X$ is an invertible measurable transformation preserving μ . In many cases, μ is a probability measure, and the theory of probability-preserving dynamical systems has considerably developed since the mid-twentieth century. In the immense zoo of examples that have caught the interest of mathematicians, very interesting and fundamental families are directly issued from probability theory. The purpose of this entry is to present two of them: systems arising from Gaussian processes and systems constructed from Poisson point processes. Although these two families are of different nature, and each one has to be addressed with specific techniques, a good reason to present them in parallel is that they share striking common features.

Introduction

As we all know, Gaussian distribution plays a prominent role in probability theory and in mathematics in general. It comes with a very rich structure, interesting objects and results, that fueled intensive studies by countless mathematicians. Therefore, it comes as no surprise that Gaussian distribution had its declination in ergodic theory, in the form of stationary Gaussian processes, and received much attention over the years. Their basic ergodic properties started to be studied in the middle of the twentieth century, especially by Ito (1944), Maruyama (1949), and Fomin (1950). As we can completely control their spectral structure with a relatively simple object which is the spectral measure of the generating process, Gaussian systems have proved to be a very rich source of examples of measure-theoretic dynamical systems with specific properties (see the examples given by Girsanov (1958), Totoki (1964), and Newton (1966)).

The Poisson point process has a long history. It was first considered on the line to represent the occurrences of independent events over time. The idea of letting the random points move to model the behavior of a large number of particles (Dobrushin 1956; Doob 1953) led to study the Poisson point process on more general spaces. The first formalization of Poisson suspensions, due to Goldstein, Lebowitz, and Aizenman in 1975 (Goldstein et al. 1975), was also motivated by questions of statistical physics. Simultaneously and independently, Vershik et al. (1975) considered the action of a group of transformations on configurations of random points. This initiated the systematic study of ergodic properties of these systems, in particular by Marchat (1978), Grabinsky (1984) and Kalikow (1981).

It is worth mentioning Maruyama (1970) again who used both Poisson processes and Gaussian stationary processes as a tool to represent any infinitely divisible stationary processes and obtained results on ergodicity and mixing for those processes.

From Probabilistic Objects to Dynamical Systems

Gaussian Systems

Stationary Gaussian process. We only consider here Gaussian processes $(\xi_n)_{n \in \mathbb{Z}}$ that are *centered*: $\mathbb{E}[\xi_n] = 0$ for each n . A fundamental property of Gaussian processes is the fact that their distribution law is completely determined by the covariances $\mathbb{E}[\xi_i \xi_j]$. In particular, such a process is stationary if and only if the covariances satisfy, for all $i, j \in \mathbb{Z}$, $\mathbb{E}[\xi_i \xi_j] = \mathbb{E}[\xi_{i-j} \xi_0]$.

In the stationary case, the law of the Gaussian process is therefore completely determined by its spectral measure, which is a symmetric positive finite measure σ on $\mathbb{T}$ whose Fourier coefficients are given by (2).

Standard Gaussian systems. Conversely, given a symmetric positive finite measure σ on $\mathbb{T}$, there exists a unique (up to equality in distribution) centered stationary Gaussian process whose covariances are given by (2). (In some sense, this ability of realizing any spectral measure characterizes Gaussian processes (see section “[From Foias-Stratila to GAGs](#)”). The distribution μ_σ of this stationary Gaussian process is shift-invariant, thus we can consider the probability-preserving dynamical system

$$X_\sigma := (\mathbb{R}^{\mathbb{Z}}, \mathcal{B}(\mathbb{R}^{\mathbb{Z}}), \mu_\sigma, S)$$

where S denotes the shift map: $(x_n)_{n \in \mathbb{Z}} \mapsto (x_{n+1})_{n \in \mathbb{Z}}$.

A *standard Gaussian system* is a probability-preserving dynamical system $(X, \mathcal{B}, \mu, T)$ isomorphic to X_σ for some symmetric positive finite measure σ on $\mathbb{T}$. In other words, $(X, \mathcal{B}, \mu, T)$ is a standard Gaussian system if the sigma-algebra $\mathcal{B}$ is generated by some Gaussian process $(\xi_n)_{n \in \mathbb{Z}}$ with $\xi_n = \xi_0 \circ T^n$ for each $n \in \mathbb{Z}$. (Multiplying if necessary the process (ξ_n) by a constant, we can always assume that $\mathbb{E}[\xi_0^2] = 1$, and in this case the spectral measure σ is a probability measure.)

The classical theory of Gaussian systems studies the above-mentioned *standard* Gaussian systems. But following (Lemańczyk et al. 2000), it is useful to introduce also the *generalized Gaussian systems* as probability-preserving dynamical systems $(X, \mathcal{B}, \mu, T)$ satisfying the following more general property: there exists a closed real subspace H of $L^2(\mu)$ such that

- H is a Gaussian space.
- H is invariant by U_T (in particular, the stationary process $(\xi \circ T^n)_{n \in \mathbb{Z}}$ is Gaussian).
- The sigma-algebra generated by H is $\mathcal{B}$.

Geometric interpretation. A geometric model for a standard Gaussian dynamical system is proposed in de la Rue (1995) as a transformation of a complex Brownian motion path. More precisely, let $B = (B_t)_{0 \leq t \leq 1}$ be a complex Brownian motion with $B_0 = 0$. For any given probability measure γ on $\mathbb{T}$ (identified here with $[0, 1]$), let us define, for $0 \leq s < 1$

$$\theta(s) := \inf \{x \in [0, 1] : \gamma([0, x]) \geq s\}.$$

Now we consider a new process $\tilde{B} = (\tilde{B}_t)_{0 \leq t \leq 1}$ given by

$$\tilde{B}_t := \int_0^t e^{i2\pi\theta(s)} dB_s. \quad (3)$$

Then $\tilde{B}$ is also a complex Brownian motion with the same law as B , and this enables us to define a transformation $T_\gamma : B \mapsto \tilde{B}$ of the canonical space $C_0([0, 1], \mathbb{C})$ of the Brownian motion, preserving the Wiener measure μ_W . This transformation is invertible (its inverse being given by a similar formula, where we replace $e^{i2\pi\theta(s)}$ by $e^{-i2\pi\theta(s)}$). Then we define the real valued stationary process $(\xi_p)_{p \in \mathbb{Z}}$ by

$$\xi_p := \Re e(B_1 \circ T_\gamma^p).$$

It turns out that $(\xi_p)_{p \in \mathbb{Z}}$ is a stationary Gaussian process, whose spectral measure σ is the

symmetrized of γ , defined for each measurable subset A of $\mathbb{T}$ by

$$\sigma(A) = \frac{\gamma(A) + \gamma(-A)}{2}.$$

Moreover, if σ is continuous, $(\xi_p)_{p \in \mathbb{Z}}$ generates the same sigma-algebra as $(B_t)_{0 \leq t \leq 1}$, and the measure-preserving system $(C_0([0, 1], \mathbb{C}), \mu_W, T_\gamma)$ is a Gaussian dynamical system isomorphic to X_σ .

Poisson Suspensions

Poisson point process. Let $(X, \mathcal{A})$ be a standard Borel space, and let $(X^*, \mathcal{A}^*)$ be the canonical space of point processes on $(X, \mathcal{A})$, where

- X^* is the set of counting measures on X ,
- $\mathcal{A}^*$ is the sigma-algebra generated by the maps $(\mathcal{N}_A)_{A \in \mathcal{A}}$, where

$$\mathcal{N}_A : \begin{cases} X^* \rightarrow \mathbb{N} \cup \{+\infty\} \\ \omega \mapsto \omega(A). \end{cases}$$

When μ is a sigma-finite measure on $(X, \mathcal{A})$, a Poisson point process of intensity μ on X always exists. We denote by μ^* its distribution, which is a probability measure on $(X^*, \mathcal{A}^*)$.

Poisson suspensions. The distribution μ^* of a Poisson point process of intensity μ has many nice functorial features. One of the most important is the following.

Let μ and ν be two sigma-finite measures on the standard Borel spaces $(X, \mathcal{A})$ and $(Y, \mathcal{B})$, respectively, and consider a measurable map $\varphi : X \rightarrow Y$ such that $\mu \circ \varphi^{-1} = \nu$. Then the map $\varphi_* : X^* \rightarrow Y^*$ acting by

$$\varphi_* \left(\sum_{i \in I} \delta_{x_i} \right) := \sum_{i \in I} \delta_{\varphi(x_i)}$$

satisfies $\mu^* \circ \varphi^{-1} = \nu^*$.

In particular, if T is a measure-preserving transformation on $(X, \mathcal{A}, \mu)$, then T_* is a probability-preserving transformation of $(X^*, \mathcal{A}^*, \mu^*)$. The

system $(X^*, \mathcal{A}^*, \mu^*, T_*)$ is called the *Poisson suspension* over the base $(X, \mathcal{A}, \mu, T)$.

Of course the properties of the Poisson suspensions depend on those of the base system, and this allows to make strong connections between infinite-measure-preserving systems and probability-preserving systems.

Spectral Theory

Basics of Spectral Theory

Let us recall some basic notions of spectral theory (see e.g., (Lemańczyk 2011)).

Let U be a unitary operator acting on a separable Hilbert space H , and let $h \in H$. We denote by σ_h the spectral measure of h (see (1)). Let $C(h)$ be the *cyclic space* of h under U , that is, the closure of the linear span of the vectors $U^n h$, $n \in \mathbb{Z}$. The linear map between $C(h)$ and $L^2(\sigma_h)$ that maps $U^n h$ to $e^{i2\pi n \cdot}$ extends to an isometry and intertwines the operator U with the unitary operator

$$V : f \mapsto e^{i2\pi \cdot} f, \quad f \in L^2(\sigma_h). \quad (4)$$

The *maximal spectral type* of U is the equivalence class of $\sigma_{h_{\max}}$ for some $h_{\max}$ satisfying:

$$\forall g \in H, \quad \sigma_g \ll \sigma_{h_{\max}}.$$

(Such an $h_{\max}$ always exists.)

We say that U has *simple spectrum* whenever H itself is a cyclic space $C(h)$ for some vector h .

Whenever $(X, \mathcal{A}, \mu, T)$ is a measure-preserving dynamical system with a sigmafinite measure μ , we denote by $U_T : L^2(\mu) \rightarrow L^2(\mu)$ the associated *Koopman operator* defined by $U_T h := h \circ T$. It is a unitary operator on $L^2(\mu)$.

Fock Space

We describe here an algebraic construction that plays a crucial role in the study of our objects.

Let once again H be a Hilbert space and denote $H^{\odot n}$, the vector space of symmetric elements of the n -th tensor product $H^{\otimes n}$. When H is $L^2(X, \mu)$ for some sigma-finite measure μ on X , then $H^{\odot n}$ can be identified with the subspace $L^2_{\text{perm}}(X^n, \mu^{\otimes n})$ of

$L^2(X^n, \mu^{\otimes n})$ consisting of functions which are invariant by coordinate permutations.

With the convention $H^{\odot 0} := \mathbb{C}$, we can consider the vector space $\sum_{n \geq 0} H^{\odot n}$ which is the formal direct sum whose elements are finite sums of vectors of $H^{\odot n}$, $n \geq 1$. The space $\sum_{n \geq 0} H^{\odot n}$ can be equipped with a scalar product by considering the direct sum as orthogonal and by endowing each $H^{\odot n}$ with the scalar product $\langle \cdot, \cdot \rangle_{H^{\otimes n}}$. The (Boson) *Fock space* $F(H)$ of H is the Hilbert space obtained as the completion of $\sum_{n \geq 0} H^{\odot n}$ with respect to the norm of the scalar product we just set up.

Operators on a Fock Space

Whenever Φ is an operator on H of norm less than or equal to 1, it extends naturally to an operator $\tilde{\Phi}$ on the Fock space $F(H)$ by acting on $H^{\odot n}$ as $\Phi^{\odot n}$, that is

$$\forall v \in H, \quad \tilde{\Phi}(v \otimes \dots \otimes v) := \Phi v \otimes \dots \otimes \Phi v.$$

This operator $\tilde{\Phi}$ is called the *second quantization* of Φ (see Attal n.d., Chap. 8, p. 16).

The following proposition considers the second quantization U of a unitary operator U .

Proposition 4.1 *If U is unitary on H and $\sigma_{\max}$ is in the equivalence class of its maximal spectral type, then*

- *The second quantization $\tilde{U}$ of U is unitary on $F(H)$.*
- *The maximal spectral type of $\tilde{U}$ is the equivalence class of $\sum_{n \geq 0} \sigma_{\max}^{*n}$ (where $\sigma_{\max}^{*0} := \delta_0$).*
- *$\tilde{U}$ has simple spectrum if and only if for all n , $U^{\odot n}$ has simple spectrum and for all $n \neq m$, $\sigma_{\max}^{*n} \perp \sigma_{\max}^{*m}$.*

Application to Gaussian and Poisson Chaos

Fock Space Structure of L^2 for Gaussian Dynamical Systems and Poisson Suspensions

In the case of the standard Gaussian dynamical system X_σ , generated by the Gaussian process $(\xi_n)_{n \in \mathbb{Z}}$, we denote by H'_1 the real Gaussian

subspace of $L^2(\mu_\sigma)$ spanned by the random variables ξ_n , $n \in \mathbb{Z}$, and $H_1 := H_1^r + iH_1^i$ the complex subspace spanned by the process. Then H_1 is isometric to $L^2(\mathbb{T}, \sigma)$, with an isometry extending the correspondence $\xi_n \leftrightarrow e^{i2\pi n}$ ($n \in \mathbb{Z}$).

In the case of the Poisson suspension $(X^*, \mathcal{A}^*, \mu^*, T_*)$, we denote by H_1 the subspace of $L^2(\mu^*)$ spanned by the random variables of the form $\mathcal{N}_A - \mu(A)$, $A \in \mathcal{A}$, $\mu(A) < \infty$. In this case H_1 is isometric to $L^2(X, \mu)$, with an isometry extending the correspondence $\mathcal{N}_A - \mu(A) \leftrightarrow \mathbf{1}_A$ ($A \in \mathcal{A}$, $\mu(A) < \infty$).

In both cases, H_0 denotes the subspace of constant functions. Then for each $n \geq 2$, we define inductively the subspace H_n as the orthocomplement of $\bigotimes_{0 \leq j < n} H_j$ in the space of all polynomials of degree at most n in variables in H_1 (note that elements of H_1 always have moments of any order). In the classical probabilist terminology, the subspace H_n is called the n -th chaos.

It turns out that H_n is, in the Gaussian case, isometric to $L^2_{\text{perm}}(\mathbb{T}^n, \sigma^{\otimes n})$, whereas in the Poisson case it is isometric to $L^2_{\text{perm}}(X^n, \mu^{\otimes n})$. We therefore get the following description of L^2 as a Fock space, which in the Gaussian case takes the form

$$L^2(\mu_\sigma) = \bigoplus_{n \geq 0} H_n \simeq F(L^2(\mathbb{T}, \sigma)),$$

and in the Poisson case

$$L^2(\mu^*) = \bigoplus_{n \geq 0} H_n \simeq F(L^2(X, \mu)).$$

We refer to Cornfeld et al. (1982) for more details about the Fock space structure in the Gaussian case, and to Last and Penrose (2018) and Neretin (1996) in the Poisson case.

Second Quantization Operators

Once the Fock space structure of $L^2(\mu_\sigma)$ (in the Gaussian case) and $L^2(\mu^*)$ (in the Poisson suspension case) is established, it is natural to look at the Koopman operator that is associated to the underlying transformation. In both cases, this Koopman operator leaves each chaos H_n invariant, and is nothing but the second quantization of its restriction to H_1 . More precisely, we can state the following.

In the case of the standard Gaussian system X_σ , the action of U_S on the first chaos H_1 corresponds to the multiplication by $e^{i2\pi}$ in $L^2(\mathbb{T}, \sigma)$. In general, the action of U_S on the n -th chaos H_n corresponds to the multiplication by the function $(t_1, \dots, t_n) \mapsto e^{i2\pi(t_1 + \dots + t_n)}$ in $L^2_{\text{perm}}(\mathbb{T}^n, \sigma^{\otimes n})$.

In the case of the Poisson suspension $(X^*, \mathcal{A}^*, \mu^*, T_*)$, the action of U_{T_*} on H_1 corresponds to the action of U_T on $L^2(\mu)$. In general, the action of U_{T_*} on the n -th chaos H_n corresponds to the action of $U_{T^{\times n}}$ on $L^2_{\text{perm}}(X^n, \mu^{\otimes n})$.

Basic Ergodic Properties

Ergodicity and Mixing

Ergodic properties as spectral properties. We list some classical ergodic properties that are spectral by nature. We start by the probability-preserving case where it is customary to consider $U_T: f \mapsto f \circ T$ acting on $L^2_0(\mu) = L^2(\mu) \ominus \mathbb{C}$, the orthocomplement of the constant functions as U_T acts trivially on the latter: $U_T(a\mathbf{1}_X) = a\mathbf{1}_X$. In that case, the *reduced maximal spectral type* denotes the maximal spectral type of U_T acting on $L^2_0(\mu)$.

Theorem 5.1 *Let $(X, \mathcal{A}, \mu, T)$ be a dynamical system with $\mu(X) = 1$, U_T its Koopman operator acting on $L^2_0(\mu)$ and $\rho^0_{\max}$, a finite measure in the measure class of the reduced maximal spectral type.*

- $(X, \mathcal{A}, \mu, T)$ is ergodic if and only if $\rho^0_{\max}(\{0\}) = 0$.
- $(X, \mathcal{A}, \mu, T)$ is weakly mixing if and only if $\rho^0_{\max}$ is continuous.
- $(X, \mathcal{A}, \mu, T)$ is (strongly) mixing if and only if $\rho^0_{\max}$ is Rajchman, i.e., $\widehat{\rho^0_{\max}}(k) \xrightarrow{|k| \rightarrow \infty} 0$.

When μ is not finite, nonzero constant functions are not in $L^2(\mu)$ and that makes a radical difference.

Theorem 5.2 *Let $(X, \mathcal{A}, \mu, T)$ be a dynamical system with μ sigma-finite, U_T its Koopman*

operator acting on $L^2(\mu)$ and $\rho_{\max}$ a finite measure in the measure class of the maximal spectral type.

- $(X, \mathcal{A}, \mu, T)$ has no T -invariant set of finite positive measure if and only if $\rho_{\max}$ is continuous.
- $(X, \mathcal{A}, \mu, T)$ is of zero type (i.e., $\mu(A \cap T^n B) \rightarrow 0$ as n tends to infinity for all finite μ -measure sets A and B in $\mathcal{A}$) if and only if $\rho_{\max}$ is Rajchman.

In particular, these two properties can only happen if μ is infinite. Note also that ergodicity is no longer a spectral property in the infinite measure case.

Ergodicity and mixing of Gaussian systems and Poisson suspensions. Ergodic and mixing properties of Gaussian dynamical systems are easily characterized in terms of the spectral measure σ of the generating stationary Gaussian process. Using the spectral characterizations of ergodicity and mixing given in Theorem 5.1 and the expression of the maximal spectral type of the second quantization of a unitary operator in Proposition 4.1, we can obtain the following results which were first proved by Maruyama (1949) and Fomin (1950). Details can also be found in Cornfeld et al. (1982).

Theorem 5.3 *The Gaussian dynamical system X_σ is ergodic if and only if the spectral measure σ is continuous, that is $\sigma(\{t\}) = 0$ for each $t \in \mathbb{T}$. In this case it is also weakly mixing.*

The system X_σ is (strongly) mixing if and only if σ is Rajchman.

In particular, by the Riemann-Lebesgue Lemma, if σ is absolutely continuous, then the associated Gaussian system is mixing, which was first proved by Ito (1944). But there also exist singular spectral measures whose Fourier coefficients vanish at infinity (Menchoff 1916).

The same kinds of arguments apply to Poisson suspensions: together with Theorem 5.2, this

yields to the following characterizations (proved by Marchat (1978) and Grabinsky (1984)).

Theorem 5.4 *The Poisson suspension $(X^*, \mathcal{A}^*, \mu^*, T_*)$ is ergodic if and only if the base system $(X, \mathcal{A}, \mu, T)$ has no invariant set of finite positive measure. In this case it is also weakly mixing.*

The Poisson suspension $(X^, \mathcal{A}^*, \mu^*, T_*)$ is (strongly) mixing if and only if the base system $(X, \mathcal{A}, \mu, T)$ is of zero type.*

Leonov (1960) proved that, if a Gaussian system is mixing, then it is mixing of all orders (see also (Totoki 1964) for a simple proof by Totoki). The same result holds for a Poisson suspension (see Grabinsky 1984; Marchat 1978).

Entropy, Bernoulli Properties

The Kolmogorov-Sinai entropy of a Gaussian system is given by the following result, already stated by Pinsker in 1960 (Pinsker 1960) (see also (Lemáńczyk 1998; Newton 1971; de la Rue 1993) for various proofs).

Theorem 5.5 (Entropy of a standard Gaussian system) *Let σ be a symmetric positive finite measure on $\mathbb{T}$. If σ is singular with respect to the Lebesgue measure, then $h(X_\sigma) = 0$. Otherwise, $h(X_\sigma) = \infty$.*

In the case where its spectral measure is the Lebesgue measure λ on $\mathbb{T}$, the stationary Gaussian process $(\xi_n)_{n \in \mathbb{Z}}$ is composed of orthogonal, hence independent, random variables. Thus X_λ is a Bernoulli shift of infinite entropy. Now if $\sigma \ll \lambda$, we will see in Theorem 6.1 that X_σ is a factor of a Bernoulli shift. By Ornstein theory (Ornstein 1974), X_σ is itself isomorphic to a Bernoulli shift, and since it also has infinite entropy by Theorem 5.5, it is in fact isomorphic to X_λ .

In the general case, decomposing the spectral measure as the sum of its singular and absolutely continuous parts and assuming that both are non-zero, we get by Theorem 6.1 below that X_σ is the direct product of a Bernoulli shift of infinite

entropy with a zero-entropy Gaussian system (corresponding to the singular part of the spectral measure). In particular, any standard Gaussian system satisfies the Pinsker property (it is a product of a zero-entropy system with a Bernoulli shift).

As far as Poisson suspensions are concerned, anything can happen with the entropy which can be 0, positive finite or infinite. Together with the fact that, when the base $(X, \mathcal{A}, \mu, T)$ is a probability space, the entropy of the Poisson suspension T_* is just the same as the entropy of T , this leads to view the entropy of the Poisson suspension T_* as a possible way to define the entropy of T when T is an infinite measure-preserving transformation, as proposed in Roy (2005). Other ways to define the entropy of an infinite measure-preserving transformation, also generalizing the finite-measure case, had been proposed by Krengel (1967) and Parry (1969), and it is proved in Janvresse et al. (2010) that these three notions of entropy coincide in many cases. However, an example of a transformation with zero Krengel entropy, but whose Poisson suspension has positive entropy, is presented in Janvresse and De La Rue (2012).

Assume that there exists in the base system $(X, \mathcal{A}, \mu, T)$ a *wandering set* A (i.e., whose images $T^n A$, $n \in \mathbb{Z}$, are pairwise disjoint), and such that $X = \bigcup_{n \in \mathbb{Z}} T^n A$ (the system is called *dissipative* in this case). Then, obviously, the base is of zero type and the Poisson suspension is mixing. But thanks to the independence property of the Poisson process, the Poisson suspension is in fact Bernoulli. However, there also exist Bernoulli examples of Poisson suspensions where the base is *conservative* (i.e., there is no wandering set of positive measure). Such examples are provided by Poisson suspensions over null-recurrent Markov chains: we consider a countable set Σ , an irreducible null-recurrent stochastic matrix $P = (p_{i,j})_{i,j \in \Sigma}$, and let $q = (q_i)_{i \in \Sigma}$ be a nonzero measure which is stationary with respect to P . We can then form the associated Markov shift $(X, \mathcal{A}, \mu, T)$ where $X = \Sigma^{\mathbb{Z}}$, T is the shift transformation, and μ is the shift invariant infinite measure given on cylinder sets by the formula

$$\mu([i_0 i_1, \dots, i_k]) := q_{i_0} p_{i_0, i_1} \cdots p_{i_{k-1}, i_k}.$$

It has been proved by Kalikow (1981) and Grabinsky (1984) that the associated Poisson suspension $(X^*, \mathcal{A}^*, \mu^*, T_*)$ is Bernoulli. Moreover, as shown in Janvresse et al. (2010), its entropy is given by the following formula generalizing the entropy for positive-recurrent Markov chains:

$$h(T_*) = - \sum_{i \in \Sigma} q_i \sum_{j \in \Sigma} p_{i,j} \log p_{i,j}.$$

(Examples where the above entropy is finite are provided by Krengel (1967).)

The general structure of a Poisson suspension is, in general, not known. We can ask in particular whether there exists a Poisson suspension which is K but not Bernoulli.

Joinings, Factors, and Centralizer

Gaussian Factors and Centralizer

Gaussian factors of a Gaussian system. In the standard Gaussian system X_σ generated by the Gaussian process $(\xi_n)_{n \in \mathbb{Z}}$, let us take a nonzero element $\zeta \in H_1^r$, and set for each $n \in \mathbb{Z}$, $\zeta_n := \zeta \circ S^n \in H_1^r$. Then $(\zeta_n)_{n \in \mathbb{Z}}$ is a stationary Gaussian process, and generates a particular factor sigma-algebra of X_σ which we call a *Gaussian factor*. This factor can be more precisely described by the following analysis.

Spectral theory provides an isometry between the closed real subspace H_1^r of $L^2(\mu_\sigma)$ generated by the Gaussian process (ξ_n) and the following real subspace of $L^2(\mathbb{T}, \sigma)$:

$$\begin{aligned} & L_{\text{sym}}^2(\mathbb{T}, \sigma) \\ & := \left\{ \phi \in L^2(\mathbb{T}, \sigma) : \phi(-t) = \overline{\phi(t)} \text{ for } \sigma - \text{almost every } t \right\}. \end{aligned}$$

The element ζ of H_1^r corresponds to some function $\phi \in L_{\text{sym}}^2(\mathbb{T}, \sigma)$, and the spectral measure σ_1 of the process $(\zeta_n)_{n \in \mathbb{Z}}$ is absolutely continuous with respect to σ , with density $\frac{d\sigma_1}{d\sigma} = |\phi|^2$.

Observe conversely that any symmetric positive finite measure $\sigma_1 \ll \sigma$ can be realized in this

way, for example, by taking the real-valued function $\phi := \sqrt{\frac{d\sigma_1}{d\sigma}} \in L^2_{\text{sym}}(\mathbb{T}, \sigma)$.

The Gaussian factor generated by $(\zeta_n)_{n \in \mathbb{Z}}$ only depends on the equivalence class of σ_1 , so we can denote it by $\mathcal{F}_{\sigma_1}$. The action of S on $\mathcal{F}_{\sigma_1}$ is isomorphic to X_{σ_1} .

If σ_1 is equivalent to σ , then $\mathcal{F}_{\sigma_1}$ coincides with the Borel sigma-algebra of X_σ , and X_{σ_1} is isomorphic to X_σ . On the other hand, if σ_1 is not equivalent to σ , we can find another symmetric positive finite measure σ_2 on $\mathbb{T}$ such that

- σ_2 is singular with respect to σ_1 .
- σ is equivalent to $\sigma_1 + \sigma_2$.

Using the fact that orthogonal subspaces of H^r_1 correspond to *independent* families of Gaussian random variables, we then get another Gaussian factor $\mathcal{F}_{\sigma_2}$ which is independent of $\mathcal{F}_{\sigma_1}$, and such that $\mathcal{F}_{\sigma_1}$ and $\mathcal{F}_{\sigma_2}$ together generate the same sigma-algebra as the original Gaussian process (ξ_n) . This shows that any strict Gaussian factor $\mathcal{F}_{\sigma_1}$ of X_σ admits an independent complement. In the case when X_σ is ergodic, this independent complement is itself weakly mixing, therefore the corresponding extension $X_\sigma \rightarrow X_{\sigma_1}$ is relatively weakly mixing.

We can summarize some of the above results through the following theorem.

Theorem 6.1 *Let $\sigma, \sigma_1, \sigma_2$ be symmetric positive finite measures on $\mathbb{T}$.*

- *If σ_1 and σ_2 are equivalent, then X_{σ_1} and X_{σ_2} are isomorphic.*
- *If $\sigma_1 \ll \sigma$, then X_{σ_1} is a factor of X_σ and corresponds to a Gaussian factor $\mathcal{F}_{\sigma_1}$ of X_σ .*
- *If σ is equivalent to $\sigma_1 + \sigma_2$, where σ_1 and σ_2 are mutually singular, then X_σ is isomorphic to the direct product $X_{\sigma_1} \times X_{\sigma_2}$.*

Gaussian centralizer. The geometric interpretation of a standard Gaussian dynamical system as a transformation of the Brownian motion path given in section “[Gaussian Systems](#)” allows to show that the centralizer of a Gaussian system is

always rich. First, if in Formula (3) we replace $e^{i2\pi\theta(s)}$ by $e^{i2\pi u\theta(s)}$ for $u \in \mathbb{R}$, we get a measure-preserving $\mathbb{R}$ action $(T^u_\gamma)_{u \in \mathbb{R}}$. Hence any standard Gaussian dynamical system can be embedded in an $\mathbb{R}$ -action. Moreover each T^u_γ gives rise to a generalized Gaussian system (as the Gaussian subspace H^r_1 is stable by T^u_γ).

Observe also that if we take two probability measures γ_1 and γ_2 on $\mathbb{T}$, then T_{γ_1} and T_{γ_2} always commute. Hence the centralizer of a Gaussian dynamical system always contains transformations isomorphic to any other standard Gaussian dynamical system.

The above examples all belong to the so-called *Gaussian centralizer*, whose elements are constructed as follows. We start by defining the multiplicative group

$$G_\sigma := \left\{ \psi \in L^2_{\text{sym}}(\mathbb{T}, \sigma) : |\psi| = 1 (\sigma - \text{a.e.}) \right\}.$$

For each ψ in G_σ , the multiplication by ψ in $L^2_{\text{sym}}(\mathbb{T}, \sigma)$ is a unitary operator, which corresponds in H^r_1 to a unitary operator U_ψ . Moreover, the process $(U_\psi(\xi_n))_{n \in \mathbb{Z}}$ has the same distribution as the generating Gaussian process $(\xi_n)_{n \in \mathbb{Z}}$, hence U_ψ comes from a measure-preserving transformation, which we denote here by T_ψ , commuting with S on X_σ .

Note that U_ψ can be naturally extended to a unitary operator on the complex space H^1 . The Koopman operator associated to T_ψ is then nothing but the second quantization of U_ψ .

The set of transformations

$$C^g(S) := \{T_\psi : \psi \in G_\sigma\}$$

is a subgroup of the centralizer of S which is called the *Gaussian centralizer*.

Compact and classical factors of a Gaussian dynamical system. Newton and Parry (1966) (see also Maruyama (1967)) introduced another type of factor of the Gaussian system X_σ by considering the sigma-algebra of subsets of $\mathbb{R}^{\mathbb{Z}}$ which are invariant by the transformation $(x_n)_{n \in \mathbb{Z}} \mapsto (-x_n)_{n \in \mathbb{Z}}$ (called the *even factor*). We can see that this factor is not a Gaussian factor, as the

corresponding extension is never relatively weakly mixing.

This example belongs to another class of factors of X_σ , which are obtained in the following way. Take any compact subgroup K of G_σ , which corresponds to some compact subgroup C_K of $C^\mathcal{G}(S)$, and consider

$$\mathcal{F}_K := \{A \subset \mathbb{R}^\mathbb{Z} : TA = A \text{ for each } T \in C_K\}.$$

Then $\mathcal{F}_K$ is a factor sigma-algebra of X_σ which is called a *Gaussian-compact factor* of X_σ . Note that Newton-Parry's example corresponds to the subgroup K generated by the function identically equal to -1 on $\mathbb{T}$.

Finally, these two types of factors can be combined, and we call a *classical factor* of X_σ any Gaussian-compact factor of a Gaussian factor of X_σ .

Poisson Factors and Centralizer

There are two main ways to get natural factors of a Poisson suspension of base $(X, \mathcal{A}, \mu, T)$ which are themselves Poisson suspensions. The first one is to consider a nontrivial T invariant set $Y \subset X$ and the sigma-algebra generated by restriction of the Poisson process to Y :

$$\mathcal{A}_{|Y}^* := \sigma\{\mathcal{N}_A, A \in \mathcal{A}, A \subset Y\}.$$

The associated factor map is $\omega \mapsto \omega_{|Y}$ and sends $(X^*, \mathcal{A}^*, \mu^*, T_*)$ to

$$(Y^*, (\mathcal{A}_{|Y})^*, (\mu_{|Y})^*, (T_{|Y})^*).$$

As an immediate consequence of the independence properties of a Poisson process, we get the following result.

Proposition 6.2 *Assume there exists a non trivial T -invariant set $Y \subset X$. Then the Poisson suspension $(X^*, \mathcal{A}^*, \mu^*, T_*)$ is isomorphic to the direct product of the two Poisson suspensions*

$$\begin{aligned} & (Y^*, (\mathcal{A}_{|Y})^*, (\mu_{|Y})^*, (T_{|Y})^*) \text{ and} \\ & ((X \setminus Y)^*, (\mathcal{A}_{|X \setminus Y})^*, (\mu_{|X \setminus Y})^*, (T_{|X \setminus Y})^*). \end{aligned}$$

The other way to get a Poisson suspension as a factor of $(X^*, \mathcal{A}^*, \mu^*, T_*)$ is to consider a sigma-finite factor of the base system $C \subset \mathcal{A}$, that is, a T -invariant sigma-algebra where μ is still sigma-finite. We then define

$$C^* := \sigma\{\mathcal{N}_A, A \in C\}.$$

We can check that the action of T_* on C^* can be viewed as a Poisson suspension. Indeed, we have the isomorphism

$$\begin{aligned} & ((X/C)^*, (\mathcal{A}/C)^*, (\mu/C)^*, (T/C)^*) \\ & \simeq (X_{/C^*}^*, \mathcal{A}_{/C^*}^*, \mu_{/C^*}^*, T_{*/C^*}). \end{aligned}$$

In general, we call *Poisson factor* of the suspension $(X^*, \mathcal{A}^*, \mu^*, T_*)$ any T_* -invariant sub-sigma-algebra of the form $(C_{|Y})^*$, where C is a sigma-finite factor of the base system and Y a C -measurable T -invariant subset of X .

The centralizer of T_* always contains other Poisson suspensions, namely all transformations of the form S_* when $S \in C(T)$. They form a subgroup of $C(T_*)$: the *Poisson centralizer*.

Gaussian and Poisson Self-Joinings

This section is based on notions borrowed from Lemańczyk et al. (2000) (for Gaussian) and Roy (2009), Derriennic et al. (2008) (for Poisson). The inherent probabilistic nature of Gaussian and Poisson systems allows to define remarkable families of self-joinings.

In the Gaussian setting, the definition is pretty straightforward.

Definition 6.3 *Let v be a self-joining of a standard Gaussian system X_σ . In the probability-preserving dynamical system*

$$(\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}}, \mathcal{B}(\mathbb{R}^{\mathbb{Z}}) \otimes \mathcal{B}(\mathbb{R}^{\mathbb{Z}}), \nu, S \times S),$$

the two natural projections on $\mathbb{R}^{\mathbb{Z}}$ provide two copies

$$\begin{aligned} (\xi'_n) &= (\xi'_0 \circ (S \times S)^n)_{n \in \mathbb{Z}} \text{ and } (\xi''_n) \\ &= (\xi''_0 \circ (S \times S)^n)_{n \in \mathbb{Z}} \end{aligned}$$

of the original Gaussian process. The self-joining ν is said to be a Gaussian self-joining if the real subspace of $L^2(\nu)$ spanned by the ξ'_n and the ξ''_n , $n \in \mathbb{Z}$, is a Gaussian space.

Example 6.4 The product self-joinings, and more generally the relatively independent product over a Gaussian factor, are Gaussian self-joinings. The graph self-joining associated to a transformation in the Gaussian centralizer is also a Gaussian self-joining. The ergodic components of the relatively independent product over a compact factor are Gaussian self-joinings.

For the Poisson counterpart, the aforementioned family of self-joinings is easier to introduce by expliciting their structure.

Let $(X^*, \mathcal{A}^*, \mu^*, T_*)$ be a Poisson suspension. Start with a “sub-self-joining” of the base, namely a system $(X \times X, \mathcal{A} \otimes \mathcal{A}, m, T \times T)$ where m is a $T \times T$ -invariant measure whose projections satisfy

$$m_1 := m(\cdot \times X) \leq \mu \text{ and } m_2 := m(X \times \cdot) \leq \mu.$$

We form the Poisson suspension $((X \times X)^*, (\mathcal{A} \otimes \mathcal{A})^*, m^*, (T \times T)_*)$. Consider the map $\kappa \mapsto (\omega_1, \omega_2) := (\kappa(\cdot \times X), \kappa(X \times \cdot))$, and denote by $\bar{m}$ the pushforward measure of m^* by this map. Then we get as a factor

$$(X^* \times X^*, \mathcal{A}^* \otimes \mathcal{A}^*, \bar{m}, T_* \times T_*),$$

which is nothing else than a joining between $(X^*, \mathcal{A}^*, m_1^*, T_*)$ and $(X^*, \mathcal{A}^*, m_2^*, T_*)$.

The final step consists into “fixing the intensities” by considering the direct product of the three systems

$$\begin{aligned} &(X^*, \mathcal{A}^*, (\mu - m_1)^*, T_*), \\ &(X^* \times X^*, \mathcal{A}^* \otimes \mathcal{A}^*, \bar{m}, T_* \times T_*), \\ &\text{and} \\ &(X^*, \mathcal{A}^*, (\mu - m_2)^*, T_*). \end{aligned}$$

On $X^* \times X^*$, we define $\tilde{m}$ as the pushforward measure of $(\mu - m_1)^* \otimes \bar{m} \otimes (\mu - m_2)^*$ by the factor map

$$(\omega', (\omega_1, \omega_2), \omega'') \mapsto (\omega' + \omega_1, \omega_2 + \omega'').$$

We get a self-joining $(X^* \times X^*, \mathcal{A}^* \otimes \mathcal{A}^*, \tilde{m}, T_* \times T_*)$ of the original Poisson suspension.

Definition 6.5 We call Poisson self-joining of a Poisson suspension any self-joining which is obtained as explained above.

Example 6.6 The product self-joining corresponds to the case where m is the null measure. If S is in the centralizer of T and $m := \Delta_S$ is the corresponding graph measure, then $\tilde{m} = \Delta_{S_*}$, i.e., the graph self-joining associated to S_* . If $C \subset \mathcal{A}$ is a sigma-finite factor and $m := \mu \otimes_C \mu$ is the relatively independent joining over C , then $\tilde{m} = \mu^* \otimes_C \mu^*$, i.e., the relatively independent joining over the Poisson factor C^* .

Despite the different nature of their structure, it is possible to give a unified characterization of Poisson and Gaussian self-joinings (see Roy 2005).

Proposition 6.7 A self-joining of a Gaussian system (resp. a Poisson suspension) is Gaussian (resp. Poisson) if and only if its distribution is infinitely divisible.

In the Gaussian case, this follows immediately from the fact that a process with Gaussian entries is Gaussian if and only if it is infinitely divisible.

Theorem 6.8 (Derriennic et al. 2008; Lemańczyk et al. 2000) A self-joining of a Gaussian system X_σ is Gaussian if and only if the associated Markov operator Φ acting on $L^2(\mu_\sigma)$ is the second quantization of an operator φ on $L^2(\sigma)$ of norm less than or equal to one and which commutes with V (see (4)).

Similarly, a self-joining of a Poisson suspension $(X^*, \mathcal{A}^*, \mu^*, T^*)$ is Poisson if and only if the associated Markov operator Φ acting on $L^2(\mu^*)$ is the second quantization of an operator φ on $L^2(\mu)$ corresponding to a sub-self-joining of $(X, \mathcal{A}, \mu, T)$ (a sub-Markov operator commuting with T).

The following property is central in this theory (see (Lemańczyk et al. 2000) for Gaussian, (Roy 2005) and (Derriennic et al. 2008) for Poisson).

Proposition 6.9 *A Gaussian (resp. Poisson) self-joining of an ergodic Gaussian system (resp. Poisson suspension) is ergodic.*

GAGs and PAPs

From Foias-Stratila to GAGs

Lemańczyk, Parreau, and Thouvenot introduced in Lemańczyk et al. (2000) a class of Gaussian dynamical systems called GAGs (a French acronym meaning “Gaussian systems with Gaussian self-joinings”), on which they have developed a beautiful theory. The proof of the existence of GAGs relies on a striking rigidity result by Foias and Stratila (1967), according to which there are some spectral measures that, in ergodic systems, can only be realized by Gaussian processes (this result improved a previous work of Sinai (1963)). Foias-Stratila result can be formulated in the following way (see also (Cornfeld et al. 1982, p. 375)).

Theorem 7.1 (Foias-Stratila) *Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ be an ergodic probability-preserving dynamical system. Assume that for some nonzero real-valued $\xi_0 \in L^2(\mathbb{P})$, the spectral measure of the stationary process $(\xi_0 \circ T^n)_{n \in \mathbb{Z}}$ is continuous and concentrated on $K \cup (-K)$, where K is a Kronecker subset of $\mathbb{T}$. Then the process $(\xi_0 \circ T^n)_{n \in \mathbb{Z}}$ is Gaussian.*

The above result is used in Lemańczyk et al. (2000) through the following argument (which was already stated by Thouvenot (1987, 1996)). Let us consider a Gaussian-Kronecker system X_σ

(which means that the spectral measure σ satisfies the same assumptions as in Theorem 7.1). Let ν be a self-joining of X_σ , and denote by ξ' and ξ'' the two copies of the original Gaussian process with spectral measure σ in the probability-preserving dynamical system

$$(\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}}, \mathcal{B}(\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}}), \nu, S \times S),$$

(as in Definition 6.3). In $L^2(\nu)$, let H' and H'' be the real subspaces spanned, respectively, by ξ' and ξ'' , and $H := H' + H''$. For any $h \in H$, the spectral measure of the stationary process $(h \circ (S \times S)^n)_{n \in \mathbb{Z}}$ is absolutely continuous with respect to σ . Hence, if we further assume that ν is ergodic, Theorem 7.1 implies that H is a Gaussian subspace.

We formalize the above by the following definition and theorem.

Definition 7.2 *The Gaussian system X_σ is said to be a GAG when any ergodic self-joining of X_σ is a Gaussian self-joining.*

Note that in this case, the system given by any ergodic self-joining is a generalized Gaussian system.

Theorem 7.3 *Any Gaussian-Kronecker system is a GAG.*

Even if Foias-Stratila result is the keystone for the construction of GAGs, it is shown in Lemańczyk et al. (2000) that the class of GAGs extends far beyond the Gaussian-Kronecker systems. Indeed, it is proved that any simple-spectrum Gaussian system is GAG, and in particular there exist *mixing* GAGs.

Because of their special properties (see section “Properties of GAGs and PAPs”), GAGs can be used to produce very interesting examples of probability-preserving dynamical systems. But beside the theory of GAGs, Foias-Stratila property can lead to other surprising results. For example, it is used in de la Rue (1998a) to show the existence of some spectral measures on the torus that can never be obtained in a transformation induced by an irrational rotation of the circle.

Poissonian Analog of Foias-Stratila Theorem and PAPs

Roughly speaking, an analog of Foias-Stratila Theorem in the context of Poisson suspensions should identify a special class of infinite measure-preserving dynamical systems $(X, \mathcal{A}, \mu, T)$ playing a role similar to the Kronecker measures, such that any *point* process invariant and ergodic under the dynamics of T would have a Poissonian character.

This invariance requirement is encompassed in the following notion of T -point process.

Definition 7.4 *Let $(X, \mathcal{A}, \mu, T)$ be a sigma-finite measure-preserving system, and let $(\Omega, \mathbb{F}, \mathbb{P}, S)$ be a probability-preserving system. A point process $N : \Omega \rightarrow X^*$ is said to be a T -point process if*

- for any $A \in \mathcal{A} : N(A) \circ S = N(T^{-1}A)$
- $N(A) = 0 \mathbb{P}$ -a.s. if $\mu(A) = 0$

The T -point process N is T -free if its distribution is concentrated on the set of simple counting measures of the form $\sum_{i \in I} \delta_{x_i}$ such that for each $i \neq j$, x_i and x_j are not in the same T -orbit.

We assume now that X is a complete and separable metric space, and that μ is an infinite and boundedly finite measure on $(X, \mathcal{A})$, that is, a measure giving finite mass to any bounded measurable set. (This definition depends on the choice of the metric, and our assumption implies in particular that the metric is unbounded.)

We consider the family (FS) of μ -preserving invertible transformations T of X satisfying:

- For each $n \geq 1$, $(X^n, \mathcal{A}^{\otimes n}, \mu^{\otimes n}, T^{\times n})$ is ergodic (T is said to have *infinite ergodic index*).
- For each $n \geq 1$, if ν is a boundedly finite $T^{\times n}$ -invariant measure on $(X^n, \mathcal{A}^n)$ whose marginals are absolutely continuous with respect to μ , then $(X^n, \mathcal{A}^{\otimes n}, \nu, T^{\times n})$ is conservative and any of its ergodic component m is of the form

$$m(A_1 \times \cdots \times A_n) = \prod_{P \in \pi} \mu \left(\bigcap_{i \in P} T^{-k_i} A_i \right),$$

where π is a partition of $\{1, \dots, n\}$ and $\{k_i\}$ are integers.

Observe that such an ergodic component gives rise to a system isomorphic to $(X^{\neq \pi}, \mathcal{A}^{\otimes \# \pi}, \mu^{\otimes \# \pi}, T^{\times \# \pi})$.

This above property is reminiscent to the concept of minimal self-joinings in the probability-preserving case. However, we stress that, when the measure is infinite, greater care is required when dealing with invariant measures on Cartesian products (see (Janvresse et al. 2017a) where this definition is discussed, and the construction of an example of transformation in the family (FS) is given).

We can now state the Poissonian version of Theorem 7.1, proved in Janvresse et al. (2017b).

Theorem 7.5 (Poissonian Analog of Foias-Stratila Theorem) *Let N be a T -point process with moments of all orders defined on the ergodic probability-preserving dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, S)$. If T is in the family (FS) and if N is T -free, then N is a Poisson point process of intensity $\alpha \mu$ for some $\alpha > 0$.*

By analogy with GAG, we can formulate the following definition.

Definition 7.6 *The Poisson suspension $(X^*, \mathcal{A}^*, \mu^{**}, T_*)$ is said to be a PAP when all its ergodic self-joinings are Poisson self-joinings.*

The analog of Theorem 7.3 in the context of Poisson suspensions has been established in Janvresse et al. (2017b).

Theorem 7.7 *If T is in the family (FS), then $(X^*, \mathcal{A}^*, \mu^*, T_*)$ is a PAP.*

Properties of GAGs and PAPs

GAGs and PAPs have the great advantage that we can quite well describe the structure of their self-

joinings. Since these self-joinings can tell a lot on the factors and centralizer of the corresponding transformation, it is no surprise that we are able to control the factors and centralizer of GAGs and PAPs.

We have seen in section “[Gaussian Factors and Centralizer](#)” that in the centralizer of any Gaussian system we can find transformations isomorphic to any other Gaussian system. It turns out that for GAGs, there is no other element in the centralizer.

Theorem 7.8 (Centralizer of a GAG (Lemańczyk et al. 2000)) *Let X_σ be a GAG, and let $T \in C(S)$. Then for each ζ in the real Gaussian subspace H_1^r of $L^2(\mu_\sigma)$ spanned by the generating Gaussian process, we have $\zeta \circ T \in H_1^r$. In particular the system $(\mathbb{R}^\mathbb{Z}, \mathcal{B}(\mathbb{R}^\mathbb{Z}), \mu_\sigma, T)$ is a generalized Gaussian system.*

We also identified in section “[Gaussian Factors and Centralizer](#)” a family of factors appearing in any standard Gaussian system. Again, in the case of GAGs, there is no other possible factor.

Theorem 7.9 (Factors of a GAG (Lemańczyk et al. 2000)) *Let X_σ be a GAG. Any factor of X_σ is a classical factor. Any factor over which S is relatively weakly mixing is a Gaussian factor.*

The above theorem is used in Iwanik et al. (1997): in the case of a GAG system, the factor generated by a function of finitely many coordinates of the Gaussian process can only be the whole system or the even factor (and the latter occurs only when the function is even).

Here are the corresponding results for Poisson suspensions.

Theorem 7.10 (Centralizer of a PAP (Janvresse et al. 2017b)) *Let $(X^*, \mathcal{A}^*, \mu^*, T_*)$ be a PAP, and let $R \in C(T_*)$. Then there exists $S \in C(T)$ such that $R = S_*$.*

In particular, when T is in the family (FS), $C(T)$ is reduced to the powers of T (Janvresse et al. 2018), thus $C(T_*) = \{T_*^n : n \in \mathbb{Z}\}$.

Theorem 7.11 (Factors of a PAP (Janvresse et al. 2017b)) *Any nontrivial factor of the PAP $(X^*, \mathcal{A}^*, \mu^*, T_*)$ contains a nontrivial Poisson factor. Any factor over which T_* is relatively weakly mixing is a Poisson factor.*

In particular, when T is in the family (FS), T_* has no nontrivial factor, as T itself has no nontrivial factor (Janvresse et al. 2018). Note that the primeness and the triviality of the centralizer of T_* in this case are striking differences compared to Gaussian systems, which always possess a lot of factors and a large centralizer. The dissemblance can be pushed even further as it is proved (see (Janvresse et al. 2017b) and (Janvresse et al. 2020)) that if T is in the family (FS), T_* is in fact disjoint from any standard Gaussian system.

Future Directions

A lot of questions remain open concerning the intersections of the classes of Gaussian systems or Poisson suspensions with other families of dynamical systems. In positive entropy, the situation is quite clear concerning the Gaussian case, but as already mentioned in section “[Entropy, Bernoulli Properties](#)” it is not known whether we can build a Poisson suspension which is K but not Bernoulli. In the zero-entropy setting, some results have been obtained for Gaussian systems: some are loosely Bernoulli, some are not (de la Rue 1996), and we know that they cannot be of finite rank (de la Rue 1998b). Similar issues for Poisson suspensions have not been studied yet. A general problem that arises is the question of disjointness of these systems of probabilistic origin with other classes such as finite-rank systems. One can hope for a unified answer taking advantage of the common features shared by Gaussian systems and Poisson suspensions, and inspired by ideas such as those developed in Derriennic et al. (2008) or Lemańczyk et al. (2011), where it is proved that both Gaussian systems and Poisson suspensions are disjoint from the class of distally simple systems.

For Poisson suspensions, the entropy theory is much richer and the situation where all three notions of entropy of a sigma-finite measure-preserving system do not coincide remains rather mysterious. For example, is it true that the Poisson entropy always dominates the Krengel entropy of the system?

It is noteworthy to mention that Poisson suspensions can also be defined over non-singular transformations, provided some integrability condition is satisfied. This is a potentially rich area to explore. (There is an ongoing work on this topic by A. Danilenko, Z. Kosloff, and the second author of this entry.)

Finally, we point out that this presentation focused on $\mathbb{Z}$ -actions, even though the notions of Gaussian systems and Poisson suspensions extend naturally to more general group actions. For flows ($\mathbb{R}$ -actions), Gaussian systems and Poisson suspensions provide interesting examples of flows for which the self-similarity set

$$I(T) := \{s \in \mathbb{R} : (T^{st})_{t \in \mathbb{R}} \text{ is isomorphic to } (T^t)_{t \in \mathbb{R}}\}$$

can be fully described (see (Danilenko and Ryzhikov 2012; Fraczek and Lemańczyk 2009; Fraczek et al. 2013)). For other groups, the generalization of some topics presented in the present survey is not always obvious. The entropy of Gaussian actions of Abelian groups is proved to be 0 or ∞ in Lemańczyk (1998); however, it is unknown if the same holds for countable amenable groups. We can also ask for which group actions we have a Foias-Stratila theory.

Bibliography

- Attal S. Lectures in quantum noise theory. <http://math.univ-lyon1.fr/attal/chapters.html>
- Cornfeld IP, Fomin SV, Sinai YG (1982) Ergodic theory. Grundlehren der Mathematischen Wissenschaften [Fundamental principles of mathematical sciences], vol 245. Springer, New York
- Danilenko AI, Ryzhikov VV (2012) On self-similarities of ergodic flows. *Proc Lond Math Soc* 104(3):431–454. (English)
- de la Rue T (1993) Entropie d'un système dynamique gaussien: cas d'une action de $\mathbb{Z}^d$. *C R Acad Sci Paris Sér I* 317(2):191–194. (French)
- de la Rue T (1995) Mouvement moyen et système dynamique gaussien. *Probab Theory Relat Fields* 102(1):45–56. (French)
- de la Rue T (1996) Systèmes dynamiques gaussiens d'entropie nulle, lâchement et non lâchement Bernoulli. *Ergodic Theory Dyn Syst* 16(2):379–404. (French)
- de la Rue T (1998a) L'induction ne donne pas toutes les mesures spectrales. *Ergodic Theory Dyn Syst* 18(6):1447–1466. (French)
- de la Rue T (1998b) Rang des systèmes dynamiques gaussiens. *Isr J Math* 104:261–283. (French)
- Derriennic Y, Fraczek K, Lemańczyk M, Parreau F (2008) Ergodic automorphisms whose weak closure of off-diagonal measures consists of ergodic self-joinings. *Colloq Math* 110(1):81–115. (English)
- Dobrushin RL (1956) On the Poisson law for the distribution of particles in space. *Ukr Mat Zh* 8:127–134. (Russian)
- Doob JL (1953) Stochastic processes. Wiley, New York. 654 S. 1953 (English)
- Foias C, Stratila S (1967) Ensembles de Kronecker dans la Théorie ergodique. *C R Acad Sci Paris* 267:166–168
- Fomin S (1950) On dynamical systems in a space of functions. *Ukr Mat Zh* 2(2):25–47. (Russian)
- Fraczek K, Lemańczyk M (2009) On the self-similarity problem for ergodic flows. *Proc Lond Math Soc* 99(3):658–696. (English)
- Fraczek K, Kulaga J, Lemańczyk M (2013) On the self-similarity problem for Gaussian-Kronecker flows. *Proc Am Math Soc* 141(12):4275–4291. (English)
- Girsanov IV (1958) Spectra of dynamical systems generated by stationary Gaussian processes. *Dokl Akad Nauk SSSR* 119:851–853. (Russian)
- Goldstein S, Lebowitz JL, Aizenman M (1975) Ergodic properties of infinite systems. *Dyn Syst Theory Appl Battelle Seattle 1974 Renc, Lect Notes Phys* 38:112–143
- Grabinsky G (1984) Poisson process over σ -finite Markov chains. *Pac J Math* 111:301–315. (English)
- Ito K (1944) On the ergodicity of a certain stationary process. *Proc Imp Acad Tokyo* 20:54–55. (English)
- Iwanik A, Lemańczyk M, de la Rue T, de Sam Lazaro J (1997) Quelques remarques sur les facteurs des systèmes dynamiques gaussiens. *Stud Math* 125(3):247–254. (French)
- Janvresse É, De La Rue T (2012) Zero Krengel entropy does not kill Poisson entropy. *Ann Inst Henri Poincaré, Probab Stat* 48(2):368–376. (English)
- Janvresse É, Meyerovitch T, Roy E, de la Rue T (2010) Poisson suspensions and entropy for infinite transformations. *Trans Am Math Soc* 362(6):3069–3094. (English)
- Janvresse É, Roy E, de la Rue T (2017a) Nearly finite chacon transformation, hal-01586869. *Annales Henri Lebesgue* 2(2019):369–414
- Janvresse É, Roy E, de la Rue T (2017b) Poisson suspensions and SuShis. *Ann Scient Éc Norm Sup* 50(6):1301–1334
- Janvresse É, Roy E, de la Rue T (2018) Invariant measures for Cartesian powers of Chacon infinite transformation. *Israel J Math* 224:1–37

- Janvresse É, Roy E, de la Rue T (2020) Ergodic poisson splittings, preprint. *Ann Probab* 48(3):1266–1285
- Kalikow S (1981) A Poisson random walk is Bernoulli. *Commun Math Phys* 81:495–499. (English)
- Krengel U (1967) Entropy of conservative transformations. *Z Wahrscheinlichkeitstheorie Verw Gebiete* 7:161–181. MR MR0218522 (36 #1608)
- Last G, Penrose M (2018) *Lectures on the Poisson process*, vol 7. Cambridge University Press, Cambridge. (English)
- Lemńczyk M (1998) Entropy of Gaussian actions for countable Abelian groups. *Fundam Math* 157(2–3):277–286. (English)
- Lemańczyk M (2011) *Spectral theory of dynamical systems*. Springer New York, New York, pp 1618–1638
- Lemańczyk M, Parreau F, Thouvenot J-P (2000) Gaussian automorphisms whose ergodic self-joinings are Gaussian. *Fundam Math* 164(3):253–293
- Lemańczyk M, Parreau F, Roy E (2011) Joining primeness and disjointness from infinitely divisible systems. *Proc Am Math Soc* 139(1):185–199. (English)
- Leonov VP (1960) The use of the characteristic functional and semi-invariants in the ergodic theory of stationary processes. *Sov Math Dokl* 1:878–881. (English)
- Marchat FA (1978) A class of measure-preserving transformations arising by the Poisson process, Ph. D. Thesis, Berkeley, Dec 1978
- Maruyama G (1949) The harmonic analysis of stationary stochastic processes. *Mem Fac Sci Kyūsyū Univ Ser A* 4:45–106
- Maruyama G (1967) A singular flow with countable Lebesgue spectrum. *J Math Soc Japan* 19:359–365. (English)
- Maruyama G (1970) Infinitely divisible processes. *Theory Probab Appl* 15(1):1–22
- Menchoff D (1916) Sur l'unicité du développement trigonométrique. *C R Acad Sci Paris* 163:433–436. (French)
- Neretin YA (1996) Categories of symmetries and infinite-dimensional groups. Transl. from the Russian by G. G. Gould. Clarendon Press, Oxford. (English)
- Newton D (1966) On Gaussian processes with simple spectrum. *Z Wahrscheinlichkeitstheor Verw Geb* 5:207–209. (English)
- Newton D (1971) Coalescence and spectrum of automorphisms of a Lebesgue space. *Z Wahrscheinlichkeitstheor Verw Geb* 19:117–122. (English)
- Newton D, Parry W (1966) On a factor automorphism of a normal dynamical system. *Ann Math Stat* 37:1528–1533. (English)
- Ornstein DS (1974) *Ergodic theory, randomness, and dynamical systems*. Yale University Press, New Haven. (English)
- Parry W (1969) Entropy and generators in ergodic theory. W. A. Benjamin, Inc., New York-Amsterdam. MR MR0262464 (41 #7071)
- Pinsker MS (1960) Dynamical systems with completely positive or zero entropy. *Sov Math Dokl* 1:937–938. (English)
- Roy E (2005) Mesures de poisson, infinite divisibilité et propriétés ergodiques. PhD thesis
- Roy E (2009) Poisson suspensions and infinite ergodic theory. *Ergodic Theory Dyn Syst* 29(2):667–683
- Sinai YG (1963) On higher order spectral measures of ergodic stationary processes. *Theory Probab Appl* 8:429–436. (English)
- Thouvenot J-P (1987) The metrical structure of some Gaussian processes, Ergodic theory and related topics II. In: *Proceedings of the conference on Georgenthal/GDR 1986*, Teubner-Texte Math vol 94, pp 195–198
- Thouvenot J-P (1996) Utilisation des processus gaussiens en théorie ergodique., *Hommage à P. A. Meyer et J. Neveu*. Société Mathématique de France, Paris, pp 303–308. (French)
- Totoki H (1964) The mixing property of Gaussian flows. *Mem Fac Sci Kyushu Univ Ser A* 18:136–139. (English)
- Vershik AM, Gel'fand IM, Graev MI (1975) Representations of the group of diffeomorphisms. *Russ Math Surv* 30(6):1–50. (English)



Ergodic Theory: Nonsingular Transformations

Alexandre I. Danilenko¹ and Cesar E. Silva²

¹B.Verkin Institute for Low Temperature Physics and Engineering of the NAS of Ukraine, Kharkiv, Ukraine

²Department of Mathematics, Williams College, Williamstown, MA, USA

Article Outline

Glossary
 Definition of the Subject
 Introduction and Basic Results
 Panorama of Examples
 Topological Groups $\text{AUT}(X, \mu)$, $\text{AUT}_2(X, \mu)$ and $\text{AUT}_1(X, \mu)$
 Orbit Theory
 Mixing Notions and Multiple Recurrence
 Dynamical Properties of IDPFT Systems
 Dynamical Properties of Nonsingular Bernoulli and Markov Shifts
 Dynamical Properties of Nonsingular Poisson Suspensions and Nonsingular Gaussian Transformations
 Spectral Theory for Nonsingular Systems
 Entropy and Other Invariants
 Nonsingular Joinings and Factors
 Smooth Nonsingular Transformations
 Miscellaneous Topics
 Applications. Connections with Other Fields
 Further Directions
 Bibliography

Glossary

Conservativity T is *conservative* if for all sets A of positive measure there exists an integer $n > 0$ such that $\mu(A \cap T^{-n}A) > 0$.

Ergodicity T is *ergodic* if every measurable subset A of X that is invariant under T (i.e., $T^{-1}A = A$) is either μ -null or μ -conull.

Equivalently, every Borel function $f: X \rightarrow \mathbb{R}$ such that $f \circ T = f$ is constant a.e.

Nonsingular dynamical system Let $(X, \mathcal{B}, \mu)$ be a standard Borel space equipped with a σ -finite measure. A Borel map $T: X \rightarrow X$ is a *nonsingular transformation* of X if for any $N \in \mathcal{B}$, $\mu(T^{-1}N) = 0$ if and only if $\mu(N) = 0$. In this case, the measure μ is called *quasi-invariant* for T , and the quadruple $(X, \mathcal{B}, \mu, T)$ is called a *nonsingular dynamical system*. If $\mu(A) = \mu(T^{-1}A)$ for all $A \in \mathcal{B}$, then μ is said to be *invariant* under T or, equivalently, T is *measure-preserving*.

Types II, II₁, II_∞, and III Suppose that μ is nonatomic and T is invertible and ergodic (and hence conservative). If there exists a σ -finite measure ν on $\mathcal{B}$ which is equivalent to μ and invariant under T then T is said to be of *type II*. It is easy to see that ν is unique up to scaling. If ν is finite, then T is of *type II₁*. If ν is infinite, then T is of *type II_∞*. If T is not of *type II*, then T is said to be of *type III*.

Definition of the Subject

An abstract measurable dynamical system consists of a set X (phase space) with a transformation $T: X \rightarrow X$ (evolution law or time) and a finite or σ -finite measure μ on X that specifies a class of negligible subsets. Nonsingular ergodic theory studies systems where T respects μ in a weak sense: The transformation preserves only the class of negligible subsets, but it may not preserve μ . This survey is about dynamics and invariants of nonsingular systems. Such systems model “non-equilibrium” situations in which events that are impossible at some time remain impossible at any other time. Of course, the first question that arises is whether it is possible to find an equivalent invariant measure, i.e., pass to a hidden equilibrium without changing the negligible subsets. It turns out that there exist systems which do not admit an equivalent invariant finite or even σ -finite measure. They are of our

primary interest here. In a way (Baire category), most of systems are like that.

Nonsingular dynamical systems arise naturally in various fields of mathematics: topological and smooth dynamics, probability theory, random walks, theory of numbers, von Neumann algebras, unitary representations of groups, mathematical physics, and so on. They also can appear in the study of probability preserving systems: some criteria of mild mixing and distality, a problem of Furstenberg on disjointness, etc. We briefly discuss this in §15. Nonsingular ergodic theory studies all of them from a general point of view:

- What is the qualitative nature of the dynamics?
- What are the orbits?
- Which properties are typical within a class of systems?
- How do we find computable invariants to compare or distinguish various systems?

Typically there are two kinds of results: Some are extensions to nonsingular systems of theorems for finite measure-preserving transformations (for instance, §2, § 4, § 12), and the other are about new properly “nonsingular” phenomena (see §5–§ 9). Philosophically speaking, the dynamics of nonsingular systems is more diverse comparatively with their finite measure-preserving counterparts. That is why it is usually easier to construct counterexamples than to develop a general theory. While infinite measure-preserving transformations are not the main subject of this survey, we cover them partially as they are also nonsingular systems and arise often as natural examples or counterexamples in the nonsingular setting. Because of shortage of space, we concentrate mainly on invertible transformations, and we have not included as many references as we had wished. General group or semigroup actions are practically not considered here (with some exceptions in §15 devoted to applications). A number of open problems are scattered through the entire text.

We thank J. Aaronson, J. R. Choksi, V. Ya. Golodets, M. Lemańczyk, F. Parreau, and E. Roy for useful remarks to the first edition of

this survey. Many new results related to nonsingular dynamical systems have appeared since the release of the first edition. The second edition is enlarged essentially to cover (partially) this progress. In particular, we added new Sections 7 and 9 and totally rewrote Section 8. More than 100 new references have been added. We are grateful to J. Aaronson, N. Avraham-Re'em, J. Hawkins, and Z. Kosloff for their remarks to the second edition of the survey.

Introduction and Basic Results

This section includes the basic results involving conservativity and ergodicity as well as some direct nonsingular counterparts of the basic machinery from classic ergodic theory: mean and pointwise ergodic theorems, Rokhlin lemma, ergodic decomposition, generators, Glimm-Effros theorem, and special representation of nonsingular flows. The historically first example of a transformation of type III (due to Ornstein) is also given here with full proof.

Nonsingular Transformations

In this survey, we will consider mainly *invertible* nonsingular transformations, i.e., those which are bijections when restricted to an invariant Borel subset of full measure. Thus when we refer to a nonsingular dynamical system $(X, \mathcal{B}, \mu, T)$, we shall assume that T is an invertible nonsingular transformation (unless the contrary is specified explicitly). Of course, each measure ν on $\mathcal{B}$ which is *equivalent* to μ , i.e., μ and ν have the same null sets, is also quasi-invariant under T . In particular, since μ is σ -finite, T admits an equivalent quasi-invariant probability measure. For each $i \in \mathbb{Z}$, we denote by ω_i^μ or ω_i the Radon-Nikodym derivative $d(\mu \circ T^i)/d\mu \in L^1(X, \mu)$. The derivatives satisfy the cocycle equation $\omega_{i+j}(x) = \omega_i(x)\omega_j(T^i x)$ for a.e. x and all $i, j \in \mathbb{Z}$.

Basic Properties of Conservativity and Ergodicity

A measurable set W is said to be *wandering* if for all $i, j \geq 0$ with $i \neq j$, $T^{-i}W \cap T^{-j}W = \emptyset$. Clearly, if T has a wandering set of positive measure, then it

cannot be conservative. A nonsingular transformation T is *incompressible* if whenever $T^{-1}C \subset C$, then $\mu(C \setminus T^{-1}C) = 0$.

Proposition 2.1 (see Krengel 1985) Let $(X, \mathcal{B}, \mu, T)$ be a nonsingular dynamical system. The following are equivalent:

- (i) T is conservative.
- (ii) For every measurable set A , $\mu(A \setminus \bigcup_{n=1}^{\infty} T^{-n}A) = 0$.
- (iii) T is incompressible.
- (iv) Every wandering set for T is null.
- (v) $\sum_{i=0}^{+\infty} \omega_i(x) = \infty$ at a.e. x (provided that $\mu(X) < \infty$).

Since any finite measure-preserving transformation is incompressible, we deduce that it is conservative. This is the statement of the classical Poincaré recurrence lemma. If T is a conservative nonsingular transformation of $(X, \mathcal{B}, \mu)$ and $A \in \mathcal{B}$ a subset of positive measure, we can define an *induced transformation* T_A of the space $(A, \mathcal{B} \cap A, \mu|_A)$ by setting $T_A x := T^n x$ if $n = n(x)$ is the smallest natural number such that $T^n x \in A$. T_A is also conservative. As shown in Silva and Thieullen (1991, 5.2), if $\mu(X) = 1$ and T is conservative and ergodic, $\int_A \sum_{i=0}^{n(x)-1} \omega_i(x) d\mu(x) = 1$, which is a nonsingular version of the well-known Kac's formula.

Theorem 2.2 (Hopf Decomposition, see Aaronson 1997). Let T be a nonsingular transformation. Then there exist disjoint invariant sets $C, D \in \mathcal{B}$ such that $X = C \sqcup D$, T restricted to C is conservative, and $D = \bigsqcup_{n=-\infty}^{\infty} T^n W$, where W is a wandering set. If $f \in L^1(X, \mu)$, $f > 0$, then $C = \{x : \sum_{i=0}^{+\infty} f(T^i x) \omega_i(x) = \infty \text{ a.e.}\}$ and $D = \{x : \sum_{i=0}^{+\infty} f(T^i x) \omega_i(x) < \infty \text{ a.e.}\}$.

The set C is called the *conservative part* of T , and D is called the *dissipative part* of T . If D is of positive measure, we call T *dissipative*. If D is of full measure, we call T *totally dissipative*.

If T is ergodic and μ is nonatomic, then T is automatically conservative. The translation by

1 on the group $\mathbb{Z}$ furnished with the counting measure is an example of an ergodic non-conservative (infinite measure-preserving) transformation.

Proposition 2.3 Let $(X, \mathcal{B}, \mu, T)$ be a nonsingular dynamical system. The following are equivalent:

- (i) T is conservative and ergodic.
- (ii) For every set A of positive measure, $\mu(X \setminus \bigcup_{n=1}^{\infty} T^{-n}A) = 0$. (In this case, we will say A sweeps out.)
- (iii) For every measurable set A of positive measure and for a.e. $x \in X$, there exists an integer $n > 0$ such that $T^n x \in A$.
- (iv) For all sets A and B of positive measure, there exists an integer $n > 0$ such that $\mu(T^{-n}A \cap B) > 0$.
- (v) If A is such that $T^{-1}A \subset A$, then $\mu(A) = 0$ or $\mu(A^c) = 0$.

A set W of positive measure is said to be *weakly wandering* if there is a sequence $n_i \rightarrow \infty$ such that $T^{n_i}W \cap T^{n_j}W = \emptyset$ for all $i \neq j$. Clearly, a finite measure-preserving transformation cannot have a weakly wandering set. Hajian and Kakutani (1964) showed that a nonsingular transformation T is of type II_1 if and only if T does not have a weakly wandering set. This entry is mainly about systems of type III . For some time, it was not quite obvious whether such systems exist at all. The historically first example was constructed by Ornstein in 1960.

Example 2.4 (Ornstein 1960) Let $A_n = \{0, 1, \dots, n\}$, $v_n(0) = 0.5$ and $v_n(i) = 1/(2n)$ for $0 < i \leq n$ and all $n \in \mathbb{N}$. Denote by (X, μ) the infinite product probability space $\otimes_{n=1}^{\infty} (A_n, v_n)$. Of course, μ is nonatomic. A point of X is an infinite sequence $x = (x_n)_{n=1}^{\infty}$ with $x_n \in A_n$ for all n . Given $a_1 \in A_1, \dots, a_n \in A_n$, we denote the cylinder $\{x = (x_i)_{i=1}^{\infty} \in X : x_1 = a_1, \dots, x_n = a_n\}$ by $[a_1, \dots, a_n]$. Define a Borel map $T : X \rightarrow X$ by setting

$$(Tx)_i = \begin{cases} 0, & \text{if } i < l(x) \\ x_i + 1, & \text{if } i = l(x) \\ x_i, & \text{if } i > l(x), \end{cases} \quad (1)$$

where $l(x)$ is the smallest number l such that $x_l \neq l$. It is easy to verify that T is a nonsingular transformation of (X, μ) and

$$\begin{aligned} \omega_1^\mu(x) &= \prod_{n=1}^{\infty} \frac{v_n((Tx)_n)}{v_n(x_n)} \\ &= \begin{cases} (l(x) - 1)!/l(x), & \text{if } x_{l(x)} = 0 \\ (l(x) - 1)!, & \text{if } x_{l(x)} \neq 0. \end{cases} \end{aligned}$$

We prove that T is of type III by contradiction. Suppose that there exists a T -invariant σ -finite measure ν equivalent to μ . Let $\varphi := d\mu/d\nu$. Then

$$\omega_i^\mu(x) = \varphi(x)\varphi(T^i x)^{-1} \text{ for a.a. } x \in X \text{ and all } i \in \mathbb{Z}. \quad (2)$$

Fix a real $C > 1$ such that the set $E_C := \varphi^{-1}([C^{-1}, C]) \subset X$ is of positive measure. By a standard approximation argument, for each sufficiently large n , there is a cylinder such that $\mu(E_C \cap [a_1, \dots, a_n]) > 0.9\mu([a_1, \dots, a_n])$. Since $v_{n+1}(0) = 0.5$, it follows that $\mu(E_C \cap [a_1, \dots, a_n, 0]) > 0.8\mu([a_1, \dots, a_n, 0])$. Moreover, by the pigeon hole principle there is $0 < i \leq n + 1$ with $\mu(E_C \cap [a_1, \dots, a_n, i]) > 0.8\mu([a_1, \dots, a_n, i])$. Find $N_n > 0$ such that $T^{N_n}[a_1, \dots, a_n, 0] = [a_1, \dots, a_n, i]$. Since $\omega_{N_n}^\mu$ is constant on $[a_1, \dots, a_n, 0]$, there is a subset $E_0 \subset E_C \cap [a_1, \dots, a_n, 0]$ of positive measure such that $T^{N_n}E_0 \subset E_C \cap [a_1, \dots, a_n, i]$. Moreover, $\omega_{N_n}^\mu(x) = v_{n+1}(i)/v_{n+1}(0) = (n+1)^{-1}$ for a.a. $x \in [a_1, \dots, a_n, 0]$. On the other hand, we deduce from (2) that $\omega_{N_n}^\mu(x) \geq C^{-2}$ for all $x \in E_0$, a contradiction.

Mean and Pointwise Ergodic Theorems. Rokhlin Lemma

Let $(X, \mathcal{B}, \mu, T)$ be a nonsingular dynamical system. Define a unitary operator U_T of $L^2(X, \mu)$ by setting

$$U_T f := \sqrt{\omega_1} \cdot f \circ T. \quad (3)$$

We note that U_T preserves the cone of positive functions $L_+^2(X, \mu)$. Conversely, every positive unitary operator in $L^2(X, \mu)$ that preserves $L_+^2(X, \mu)$ equals U_T for a μ -nonsingular transformation T . We call U_T the *Koopman operator generated by T* .

Theorem 2.5 (von Neumann mean Ergodic Theorem, see Aaronson 1997). *T has no μ -absolutely continuous T -invariant probability if and only if $n^{-1} \sum_{i=0}^{n-1} U_T^i \rightarrow 0$ in the strong operator topology.*

Proof. Let P denote the orthogonal projector in $L^2(X, \mu)$ onto the subspace of U_T -invariant vectors. By the well-known fact from the theory of Hilbert spaces, $n^{-1} \sum_{i=0}^{n-1} U_T^i \rightarrow P$ in the strong operator topology. Then $P \neq 0$ if and only if there is $f \in L^2(X, \mu)$ such that $f \neq 0$ and $U_T f = f$. Of course, $U_T |f| = |f|$. We now define a nontrivial finite measure $\lambda \ll \mu$ by setting $\frac{d\lambda}{d\mu} := |f|^2$. It is straightforward to verify that λ is invariant under T .

Denote by $\mathcal{I}$ the sub- σ -algebra of T -invariant sets. Let $\mathbb{E}_\mu[\cdot | \mathcal{I}]$ stand for the conditional expectation with respect to $\mathcal{I}$. Note that if T is ergodic, then $\mathbb{E}_\mu[f | \mathcal{I}] = \int f d\mu$. Now we state a nonsingular analogue of Birkhoff's pointwise ergodic theorem, due to Hurewicz (1944) and in the form stated by Halmos (1946).

Theorem 2.6 (Hurewicz pointwise Ergodic Theorem). *If T is conservative, $\mu(X) = 1$, $f, g \in L^1(X, \mu)$ and $g > 0$, then*

$$\frac{\sum_{i=0}^{n-1} f(T^i x) \omega_i(x)}{\sum_{i=0}^{n-1} g(T^i x) \omega_i(x)} \rightarrow \frac{\mathbb{E}_\mu[f | \mathcal{I}](x)}{\mathbb{E}_\mu[g | \mathcal{I}](x)} \text{ as } n \rightarrow \infty \text{ for a.e. } x.$$

There is a nonsingular version (Silva and Thieullen 1991) of the subadditive ergodic theorem of Kingman. Let T be a nonsingular transformation, and let (ω_n) be its sequence of Radon-Nikodym derivatives.

A sequence of functions (f_n) is said to be *subadditive* if $f_{n+m} \leq f_m + f_n \circ T^m \omega_m$ for all n ,

$m \geq 0$. Since (f_n) is a subadditive, one can verify that the following limit

$$\mathbb{S}_\mu[(f_n)](x) := \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_\mu[f_n | \mathcal{I}](x)$$

exists almost everywhere.

Theorem 2.7 (Nonsingular subadditive Ergodic Theorem). *If T is conservative, $\mu(X) = 1$, (f_n) is a subadditive sequence of integrable functions, $g \in L^1(X, \mu)$ and $g > 0$, then*

$$\frac{f_n(x)}{\sum_{i=0}^{n-1} g(T^i x) \omega_i(x)} \rightarrow \frac{\mathbb{S}_\mu[(f_n)](x)}{\mathbb{E}_\mu[g | \mathcal{I}](x)} \text{ as } n \rightarrow \infty \text{ for a.e. } x.$$

Of course, Theorem 2.6 follows from Theorem 2.7 if we set $f_n(x) := \sum_{i=0}^{n-1} f(T^i x) \omega_i(x)$ for all $n > 0$ and $x \in X$.

A transformation T is *aperiodic* if the T -orbit of a.e. point from X is infinite. The following classical statement can be deduced easily from Proposition 2.1.

Lemma 2.8 (Rokhlin's lemma (Friedman 1970)). *Let T be an aperiodic nonsingular transformation of a standard probability space (X, μ) . For each $\varepsilon > 0$ and integer $N > 1$, there exists a measurable set A such that the sets $A, TA, \dots, T^{N-1}A$ are disjoint and $\mu(A \cup TA \cup \dots \cup T^{N-1}A) > 1 - \varepsilon$.*

This lemma was refined later (for ergodic transformations) by Lehrer and Weiss as follows.

Theorem 2.9 (ϵ -free Rokhlin lemma (Lehrer and Weiss 1982)). *Let T be ergodic and μ nonatomic. Then for a subset $B \subset X$ and any N for which $\bigcup_{k=0}^{\infty} T^{-kN}(X \setminus B) = X$, there is a set A such that the sets $A, TA, \dots, T^{N-1}A$ are disjoint and $A \cup TA \cup \dots \cup T^{N-1}A \supset B$.*

The condition $\bigcup_{k=0}^{\infty} T^{-kN}(X \setminus B) = X$ holds of course for each $B \neq X$ if T is *totally ergodic*, i.e., T^p is ergodic for any p , or if N is prime. We now state a nonsingular version of Alpern's lemma which is a generalization of Lemma 2.8.

Theorem 2.10 (Alpern's lemma (Alpern and Prasad 1990)). *Let T be an aperiodic nonsingular transformation of a standard probability space (X, μ) . Let $\pi = (\pi_1, \pi_2, \dots)$ be a probability vector such that $\{k | \pi_k > 0\}$ is a relatively prime set of integers. Then there is a measurable partition $P = \{P_{k,i} | k > 0, i = 1, \dots, k\}$ of X satisfying*

- (a) $TP_{k,i} = P_{k, i+1}$ for each k and every $i < k$.
- (b) $\sum_{i=1}^k \mu(P_{k,i}) = \pi_k$ for each k .

Ergodic Decomposition

A proof of the following theorem may be found in Aaronson (1997, 2.2.8) and Aaronson (1987, §6).

Theorem 2.11 (Ergodic Decomposition Theorem). *Let T be a conservative nonsingular transformation on a standard probability space $(X, \mathcal{B}, \mu)$. There exists a standard probability space $(Y, \nu, \mathcal{A})$ and a family of probability measures μ_y on $(X, \mathcal{B})$, for $y \in Y$, such that*

- (i) *For each $A \in \mathcal{B}$, the map $y \mapsto \mu_y(A)$ is Borel and for each $A \in \mathcal{B}$*

$$\mu(A) = \int \mu_y(A) d\nu(y).$$

- (ii) *For $y, y' \in Y$, the measures μ_y and $\mu_{y'}$ are mutually singular.*
- (iii) *For each $y \in Y$, the transformation T is nonsingular and conservative, ergodic on $(X, \mathcal{B}, \mu_y)$.*
- (iv) *For each $y \in Y$, $\omega_1^{\mu_y} = \omega_1^\mu \mu_y - a.e.$*
- (v) *(Uniqueness) If there exists another probability space $(Y', \nu', \mathcal{A}')$ and a family of probability measures $\mu'_{y'}$ on $(X, \mathcal{B})$, for $y' \in Y'$, satisfying (i)–(iv), then there exists a measure-preserving isomorphism $\theta : Y \rightarrow Y'$ such that $\mu_y = \mu'_{\theta y}$ for ν -a.e. y .*

It follows that if T preserves an equivalent σ -finite measure, then the system $(X, \mathcal{B}, \mu_y, T)$ is

of type II for ν -a.a. y . The space $(Y, \nu, \mathcal{A})$ is called the *space of T -ergodic components*.

Generators

A σ -algebra $\mathcal{F}$ is called a generator for a nonsingular transformation T on a standard probability space $(X, \mathcal{B}, \mu)$, if $\bigvee_{n=-\infty}^{\infty} T^n \mathcal{F} = \mathcal{B}$. It was shown in Rokhlin (1965) and Parry (1966) that T has a *countable generator*, i.e., a countable partition P of X so that the σ -algebra of P -measurable sets is a generator for T . It was refined by Krengel (1970): If T is of type II $_{\infty}$ or III, then there exists P consisting of two sets only. Moreover, given a sub- σ -algebra $\mathcal{F} \subset \mathcal{B}$ such that $\mathcal{F} \subset T\mathcal{F}$ and $\bigcup_{k>0} T^k \mathcal{F} = \mathcal{B}$, the set $\{A \in \mathcal{F} \mid (A, X \setminus A) \text{ is a generator of } T\}$ is dense in $\mathcal{F}$. It follows, in particular, that T is isomorphic to the shift on $\{0, 1\}^{\mathbb{Z}}$ equipped with a quasi-invariant probability measure. For a version of the Krengel 2-sets generator theorem in the Borel category, we refer to Hochman (2019).

The Glimm-Effros Theorem

The classical Bogolyubov-Krylov theorem states that each homeomorphism of a compact space admits an ergodic invariant probability measure (Cornfeld et al. 1982). The following statement by Glimm (1961) and Effros (1965) is a “non-singular” analogue of that theorem. (We consider here only a particular case of $\mathbb{Z}$ -actions.)

Theorem 2.12 *Let X be a Polish space and $T : X \rightarrow X$ an aperiodic homeomorphism. Then the following are equivalent:*

- (i) T has a recurrent point x , i.e., $x = \lim_{n \rightarrow \infty} T^{n_i} x$ for a sequence $n_1 < n_2 < \dots$.
- (ii) There is an orbit of T which is not locally closed.
- (iii) There is no Borel set which intersects each orbit of T exactly once.
- (iv) There is a continuous probability Borel measure μ on X such that (X, μ, T) is an ergodic nonsingular system.

A natural question arises: Under the conditions of the theorem, how many such μ can exist? It

turns out that there is a wealth of such measures. To state a corresponding result, we first write an important definition.

Definition 2.13 Two nonsingular systems $(X, \mathcal{B}, \mu, T)$ and $(X, \mathcal{B}', \mu', T')$ are called *orbit equivalent* if there is a one-to-one bimeasurable map $\varphi : X \rightarrow X$ with $\mu' \circ \varphi \sim \mu$ and such that φ maps the T -orbit of x onto the T' -orbit of $\varphi(x)$ for a.a. $x \in X$.

The following theorem was proved in Katznelson and Weiss (1972), Krieger (1976a), and Schmidt (1977b).

Theorem 2.14 *Let (X, T) be as in Theorem 2.12. Then for each ergodic dynamical system $(Y, \mathcal{C}, \nu, S)$ of type II $_{\infty}$ or III, there exist uncountably many mutually disjoint Borel measures μ on X such that $(X, T, \mathcal{B}, \mu)$ is orbit equivalent to $(Y, \mathcal{C}, \nu, S)$.*

On the other hand, T may not have any finite invariant measure. The first such example appeared in Eigen et al. (1998b). We present a simpler one.

Example 2.15 Let T be an irrational rotation on the circle $\mathbb{T}$, and let K be a nowhere dense closed subset of $\mathbb{T}$ of positive Lebesgue measure. Let X be the complement of the T -orbit $\bigcup_{n \in \mathbb{Z}} T^n K$ of K . Then X is a T -invariant G_{δ} -subset of zero Lebesgue measure. Hence X is Polish in the induced topology and $T \upharpoonright X$ is an aperiodic homeomorphism of X . Since T is minimal, X is dense in $\mathbb{T}$ and the $(T \upharpoonright X)$ -orbit of each point of X is dense in X . Hence every point is recurrent. By Theorem 2.10, there exists a continuous ergodic nonsingular probability Borel measure λ on X . If it is invariant under $T \upharpoonright X$, then λ can be considered also as a finite T -invariant measure on $\mathbb{T}$. Since T is uniquely ergodic, λ is the Lebesgue measure. However X is of zero Lebesgue measure, a contradiction.

Let T be an aperiodic Borel transformation of a standard Borel space X . Denote by $\mathcal{M}(T)$ the set of all ergodic T -nonsingular continuous measures on X . Given $\mu \in \mathcal{M}(T)$, let $N(\mu)$ denote the family of all Borel μ -null subsets. Shelah and

Weiss showed (Shelah and Weiss 1982) that $\bigcap_{\mu \in \mathcal{M}(T)} N(\mu)$ coincides with the collection of all Borel T -wandering sets.

Minimal Radon Uniquely Ergodic Models for Infinite Measure-Preserving Transformations

We first note that there is only one, up to a homeomorphism, locally compact noncompact Cantor (i.e., zero-dimensional, perfect, and metrizable) set. Denote it by C . We recall that a Borel measure on C is called *Radon* if it is finite on every compact subset of C . The following is an infinite version of the well known Jewett-Krieger theorem.

Theorem 2.16 (Yuasa 2013). *Let T be an ergodic measure-preserving transformation of the standard infinite σ -finite measure space (X, μ) . Then there exists a minimal homeomorphism R of C that admits a unique, up to scaling, R -invariant Radon measure ν such that (X, μ, T) is isomorphic to (C, ν, R) .*

Yuasa proved a relative version of this theorem in Yuasa (2020). Similar results on *strictly ergodic models* for type III ergodic systems are not known yet. We guess that the Radon property of the measure in this case should be replaced with another property, related to the Radon-Nikodym derivative of the original transformation.

Special Representations of Ergodic Flows

Nonsingular flows ($=\mathbb{R}$ -actions) appear naturally in the study of orbit equivalence for systems of type III (see section “Orbit Theory”). Here we record some basic notions related to nonsingular flows. Let $(X, \mathcal{B}, \mu)$ be a standard Borel space with a σ -finite measure μ on $\mathcal{B}$. A nonsingular flow on (X, μ) is a Borel map $S: X \times \mathbb{R} \ni (x, t) \mapsto S_t x \in X$ such that $S_t S_s = S_{t+s}$ for all $s, t \in \mathbb{R}$ and each S_t is a nonsingular transformation of (X, μ) . Conservativity and ergodicity for flows are defined in a similar way as for transformations.

A very useful example of a flow is a flow built under a function. Let $(X, \mathcal{B}, \mu, T)$ be a nonsingular dynamical system and f a positive Borel function on X such that $\sum_{i=0}^{\infty} f(T^i x) = \sum_{i=0}^{\infty} f(T^{-i} x) = \infty$ for all $x \in X$. Set $X^f := \{(x, s) : x \in X, 0 \leq s < f(x)\}$. Define μ^f to

be the restriction of the product measure $\mu \times \text{Leb}$ on $X \times \mathbb{R}$ to X^f and define, for $t \geq 0$,

$$S_{t(x,s)}^f := \left(T^n x, s + t - \sum_{i=0}^{n-1} f(T^i x) \right),$$

where n is the unique integer that satisfies

$$\sum_{i=0}^{n-1} f(T^i x) < s + t \leq \sum_{i=0}^n f(T^i x).$$

A similar definition applies when $t < 0$. In particular, when $0 < s + t < \varphi(x)$, $S_t^f(x, s) = (x, s + t)$, so that the flow moves the point (x, s) up t units, and when it reaches $(x, \varphi(x))$, it is sent to $(Tx, 0)$. It can be shown that $S^f = (S_t^f)_{t \in \mathbb{R}}$ is a free μ^f -nonsingular flow and that it preserves μ^f if and only if T preserves μ (Nadkarni 1998). It is called the *flow built under the function ϕ with the base transformation T* . Of course, S^f is conservative or ergodic if and only if so is T .

Two flows $S = (S_t)_{t \in \mathbb{R}}$ on $(X, \mathcal{B}, \mu)$ and $V = (V_t)_{t \in \mathbb{R}}$ on $(Y, \mathcal{C}, \nu)$ are said to be *isomorphic* if there exist invariant conull sets $X' \subset X$ and $Y' \subset Y$ and an invertible nonsingular map $\rho: X' \rightarrow Y'$ that intertwines the actions of the flows: $\rho \circ S_t = V_t \circ \rho$ on X' for all t . The following nonsingular version of Ambrose–Kakutani representation theorem was proved by Krengel (1969) and Kubo (1969).

Theorem 2.17 *Let S be a free nonsingular flow. Then it is isomorphic to a flow built under a function.*

Rudolph showed that in the Ambrose–Kakutani theorem one can choose the function φ to take two values. Krengel (1976) showed that this can also be assumed in the nonsingular case.

Panorama of Examples

This section is devoted entirely to examples of nonsingular systems. We describe here the most popular (and simple) constructions of nonsingular systems: product odometers, nonsingular Markov odometers, tower transformations, rank-one finite rank systems, nonsingular Bernoulli and Markov

shifts, nonsingular Poisson suspensions and Gaussian transformations, IDPFT systems and natural extensions of nonsingular endomorphisms.

Nonsingular Product Odometers

Given a sequence m_n of natural numbers, we let $A_n := \{0, 1, \dots, m_n - 1\}$. Let v_n be a probability on A_n and $v_n(a) > 0$ for all $a \in A_n$. Consider now the infinite product probability space $(X, \mu) := \otimes_{n=1}^{\infty} (A_n, v_n)$. Assume that $\prod_{n=1}^{\infty} \max\{v_n(a) | a \in A_n\} = 0$. Then μ is nonatomic. Given $a_1 \in A_1, \dots, a_n \in A_n$, we denote by $[a_1, \dots, a_n]$ the cylinder $x = (x_i)_{i>0} | x_1 = a_1, \dots, x_n = a_n$. If $x \neq (0, 0, \dots)$, we let $l(x)$ be the smallest number l such that the l -th coordinate of x is not $m_l - 1$. We define a Borel map $T: X \rightarrow X$ by (1) if $x \neq (m_1, m_2, \dots)$ and put $Tx := (0, 0, \dots)$ if $x = (m_1, m_2, \dots)$. Of course, T is isomorphic to a rotation on a compact monothetic totally disconnected Abelian group. It is easy to check that T is μ -nonsingular and

$$\begin{aligned} \omega_1^\mu(x) &= \prod_{n=1}^{\infty} \frac{v_n((Tx)_n)}{v_n(x_n)} \\ &= \frac{v_{l(x)}(x_{l(x)} + 1)}{v_{l(x)}(x_{l(x)})} \prod_{n=1}^{l(x)-1} \frac{v_n(0)}{v_n(m_n - 1)} \end{aligned}$$

for a.a. $x = (x_n)_{n>0} \in X$. It is also easy to verify that T is ergodic. It is called the *nonsingular product odometer* associated to $(m_n, v_n)_{n=1}^{\infty}$. We note that Ornstein's transformation (Example 2.4) is a nonsingular product odometer.

Markov Odometers

We define Markov odometers as in Dooley and Hamachi (2003a). An ordered Bratteli diagram B (Herman et al. 1992) consists of

- (i) A vertex set V which is a disjoint union of finite sets $V^{(n)}$, $n \geq 0$, V_0 is a singleton.
- (ii) An edge set E which is a disjoint union of finite sets $E^{(n)}$, $n > 0$;
- (iii) Source mappings $s_n: E^{(n)} \rightarrow V^{(n-1)}$ and range mappings $r_n: E^{(n)} \rightarrow V^{(n)}$ such that $s_n^{-1}(v) \neq \emptyset$ for all $v \in V^{(n-1)}$ and $r_n^{-1}(v) \neq \emptyset$ for all $v \in V^{(n)}$, $n > 0$;

- (iv) A partial order on E so that $e, e' \in E$ are comparable if and only if $e, e' \in E^{(n)}$ for some n and $r_n(e) = r_n(e')$.

A Bratteli compactum X_B of the diagram B is the space of infinite paths

$$\left\{ x = (x_n)_{n>0} | x_n \in E^{(n)} \text{ and } r(x_n) = s(x_{n+1}) \right\}$$

on B . X_B is equipped with the natural topology induced by the product topology on $\prod_{n>0} E^{(n)}$. We will assume always that the diagram is *essentially simple*, i.e., there is only one infinite path $x_{\max} = (x_n)_{n>0}$ with x_n maximal for all n and only one $x_{\min} = (x_n)_{n>0}$ with x_n minimal for all n . The *Bratteli-Vershik* map $T_B: X_B \rightarrow X_B$ is defined as follows: $Tx_{\max} = x_{\min}$. If $x = (x_n)_{n>0} \neq x_{\max}$, then let k be the smallest number such that x_k is not maximal. Let y_k be a successor of x_k . Let $(y_1, \dots, y_k)$ be the unique path such that $y_1, \dots, y_{k-1}$ are all minimal. Then we let $T_B x := (y_1, \dots, y_k, x_{k+1}, x_{k+2}, \dots)$. It is easy to see that T_B is a homeomorphism of X_B . Suppose that we are given a sequence $P^{(n)} = (P_{(v,e)}^{(n)})_{(v,e) \in V^{n-1} \times E^{(n)}}$ of stochastic matrices, i.e.,

- (i) $P_{v,e}^{(n)} > 0$ if and only if $v = s_n(e)$.
- (ii) $\sum_{\{e \in E^{(n)} | s_n(e)=v\}} P_{v,e}^{(n)} = 1$ for each $v \in V^{(n-1)}$.

For $e_1 \in E^{(1)}, \dots, e_n \in E^{(n)}$, let $[e_1, \dots, e_n]$ denote the cylinder $\{x = (x_j)_{j>0} | x_1 = e_1, \dots, x_n = e_n\}$. Then we define a *Markov measure* on X_B by setting

$$\mu_P([e_1, \dots, e_n]) = P_{s_1(e_1), e_1}^1 P_{s_2(e_2), e_2}^2 \dots P_{s_n(e_n), e_n}^n$$

for each cylinder $[e_1, \dots, e_n]$. The dynamical system (X_B, μ_P, T_B) is called a *Markov odometer*. It is easy to see that every nonsingular product odometer is a Markov odometer where the corresponding $V^{(n)}$ are all singletons.

Tower Transformations

This construction is a discrete analogue of flow under a function. Given a nonsingular dynamical system (X, μ, T) and a measurable map $f: X \rightarrow \mathbb{N}$,

we define a new dynamical system (X^f, μ^f, T^f) by setting

$$\begin{aligned} X^f &:= \{(x, i) \in X \times \mathbb{Z}_+ \mid 0 \leq i < f(x)\}, \\ d\mu^f(x, i) &:= d\mu(x) \text{ and} \\ T^f(x, i) &:= \begin{cases} (x, i+1), & \text{if } i+1 < f(x) \\ (Tx, 0), & \text{otherwise.} \end{cases} \end{aligned}$$

Then T^f is μ^f -nonsingular and $(d\mu^f \circ T^f/d\mu^f)(x, i) = (d\mu \circ T/d\mu)(x)$ for a.a. $(x, i) \in X^f$. This transformation is called the (Kakutani) *tower over T with height function f* . It is easy to check that T^f is conservative if and only if T is conservative; T^f is ergodic if and only if T is ergodic; T^f is of type III if and only if T is of type III. Moreover, the induced transformation $(T^f)_{X \times \{0\}}$ is isomorphic to T . Given a subset $A \subset X$ of positive measure, T is the tower over the induced transformation T_A with the first return time to A as the height function.

Rank-One Transformations. Chacón Maps. Finite Rank

The definition uses the process of “cutting and stacking.” We construct by induction a sequence of columns C_n . A column C_n consists of a finite sequence of bounded intervals (left-closed, right-open) $C_n = \{I_{n,0}, \dots, I_{n,h_n-1}\}$ of height h_n . A column C_n determines a *column map* T_{C_n} that sends each interval $I_{n,i}$ to the interval above it $I_{n,i+1}$ by the unique orientation-preserving affine map between the intervals. T_{C_n} remains undefined on the top interval I_{n,h_n-1} . Set $C_0 = \{[0, 1)\}$ and let $\{r_n \geq 2\}$ be a sequence of positive integers, let $\{s_n\}$ be a sequence of functions $s_n : \{0, \dots, r_n - 1\} \rightarrow \mathbb{N}_0$, and let $\{w_n\}$ be a sequence of probability vectors on $\{0, \dots, r_n - 1\}$. If C_n has been defined, column C_{n+1} is defined as follows. First “cut” (i.e., subdivide) each interval $I_{n,i}$ in C_n into r_n subintervals $I_{n,i}[j], j = 0, \dots, r_n - 1$, whose lengths are in the proportions $w_n(0) : w_n(1) : \dots : w_n(r_n - 1)$. Next place, for each $j = 0, \dots, r_n - 1$, $s_n(j)$ new subintervals above $I_{n,h_n-1}[j]$, all of the same length as $I_{n,h_n-1}[j]$. Denote these intervals, called *spacers*, by $S_{n,0}[j], \dots, S_{n,s_n(j)-1}[j]$. This yields, for each $j \in \{0, \dots, r_n - 1\}$, r_n subcolumns each consisting of the subintervals

$$I_{n,0}[j], \dots, I_{n,h_n-1}[j] \text{ followed by the spacers } S_{n,0}[j], \dots, S_{n,s_n(j)-1}[j].$$

Finally, each subcolumn is stacked from left to right so that the top subinterval in subcolumn j is sent to the bottom subinterval in subcolumn $j+1$, for $j = 0, \dots, r_n - 2$ (by the unique orientation-preserving affine map between the intervals). For example, $S_{n,s_n(0)-1}[0]$ is sent to $I_{n,0}[1]$. This defines a new column C_{n+1} and new column map $T_{C_{n+1}}$, which remains undefined on its top subinterval. Let X be the union of all intervals in all columns, and let μ be Lebesgue measure restricted to X . We assume that as $n \rightarrow \infty$ the maximal length of the intervals in C_n converges to 0, so we may define a transformation T of (X, μ) by $Tx := \lim_{n \rightarrow \infty} T_{C_n}x$. One can verify that T is well-defined a.e. and that it is nonsingular and ergodic. T is said to be the *rank-one* transformation associated with $(r_n, w_n, s_n)_{n=1}^\infty$. If all the probability vectors w_n are uniform, the resulting transformation is measure-preserving. The measure is infinite (σ -finite) if and only if the total mass of the spacers is infinite. In the case $r_n = 3$ and $s_n(0) = s_n(2) = 0, s_n(1) = 1$ for all $n \geq 0$, the associated rank-one transformation is called a *nonsingular Chacón map*.

It is easy to see that every nonsingular product odometer is of rank-one (the corresponding maps s_n are all trivial). Each rank-one map T is a tower over a nonsingular product odometer (to obtain such an odometer, reduce T to a column C_n).

A rank N transformation is defined in a similar way. A nonsingular transformation T is said to be of *rank N or less* if at each stage of its construction there exit N disjoint columns, the levels of the columns generate the σ -algebra, and the Radon-Nikodym derivative of T is constant on each nontop level of every column. T is said to be of *rank N* if it is of rank N or less and not of rank $N-1$ or less. A rank N transformation, $N \geq 2$, need not be ergodic.

Nonsingular Bernoulli Shifts

A *nonsingular Bernoulli* transformation is a transformation T such that there exists a generating σ -algebra $\mathcal{P}$ for T (see §2.5) such that the σ -algebras, $T^n \mathcal{P}, n \in \mathbb{Z}$, are mutually independent. Thus, we may think that T is the left shift on the

probability space $(X, \mu) = (A^{\mathbb{Z}}, \otimes_{n \in \mathbb{Z}} \mu_n)$, where A is a standard Borel space and $(\mu_n)_{n \in \mathbb{Z}}$ is a sequence of probability measures on A . We will always assume that μ is nonatomic. It follows from Kakutani's criterion for equivalence of infinite product measures (Kakutani 1948) that μ is nonsingular if and only if $\mu_n \sim \mu_{n+1}$ for each n and

$$\sum_{n \in \mathbb{Z}} H^2(\mu_n, \mu_{n+1}) < \infty, \quad (4)$$

where $H^2(\mu, \nu)$ denotes the *Hellinger distance* defined by $H^2(\mu, \nu) := 1 - \int_X \sqrt{\frac{d\mu}{d\nu}} \frac{d\nu}{d\mu} d\zeta$, where ζ is a probability such that $\mu \prec \zeta$ and $\nu \prec \zeta$. If (4) is satisfied, then for a.e. $x = (x_n)_{n \in \mathbb{Z}} \in X$,

$$\omega_1^\mu(x) = \prod_{n \in \mathbb{Z}} \frac{\mu_{n-1}(x_n)}{\mu_n(x_n)}.$$

Nonsingular Markov Shifts

Let A be a finite set and let $M = (M(a, b))_{a, b \in A}$ be a 0–1-valued $A \times A$ -matrix. We let $X_M := \{x = (x_i)_{i \in \mathbb{Z}} \in A^{\mathbb{Z}} \mid M(x_i, x_{i+1}) > 0 \text{ for all } i \in \mathbb{Z}\}$. Let T denote the restriction of the left shift to X_M . Then T is a shift of finite type. Given two integers $i \leq j$, and a finite sequence $a = (a_l)_{l=i}^j$ of elements from A such that $M(a_l, a_{l+1}) = 1$ for $l = i, \dots, j-1$, we define by $[a]_i^j$ the cylinder $\{x \in X_M \mid x_l = a_l \text{ for all } l = i, \dots, j\}$. Suppose that there is a sequence of probability measures $(\pi_n)_{n \in \mathbb{Z}}$ on A and a sequence $(P_n)_{n \in \mathbb{Z}}$ of row-stochastic $A \times A$ -matrices such that $\pi_n P_n = \pi_{n+1}$ and $P_n(a, b) > 0$ if and only if $M(a, b) > 0$ for each $n \in \mathbb{Z}$. Then there is a unique probability measure μ on X_M such that for every cylinder $[a]_i^j$ in X_M .

$$\mu([a]_i^j) = \pi_i(a_i) P_i(a_i, a_{i+1}) \cdots P_{j-1}(a_{j-1}, a_j).$$

It is called Markov measure on X_M generated by $(\pi_n, P_n)_{n \in \mathbb{Z}}$. By analogy with Kakutani (1948), using Kabanov et al. (1977) one can find necessary and sufficient conditions for T to be μ -nonsingular. Then the system (X_M, T, μ) is called a *nonsingular Markov shift*. It is a natural generalization of nonsingular Bernoulli shifts. In the case where M is irreducible and $\inf\{P_n(a, b) \mid$

$M(a, b) = 1, n \in \mathbb{Z}\} > 0$, a convenient (computable) criterion for nonsingularity was found in Avraham-Re'em (2022, Theorem 5.2). By a standard computation,

$$\omega_1^\mu(x) = \lim_{n \rightarrow +\infty} \frac{\pi_{-n-1}(x_{-n})}{\pi_{-n}(x_{-n})} \prod_{j=-n}^n \frac{P_{j-1}(x_j, x_{j+1})}{P_j(x_j, x_{j+1})}.$$

We also mention here so-called *infinite Markov shifts*, i.e., Markov transformations preserving an infinite σ -finite measure (see Kakutani and Parry (1963) and §4.5 from Aaronson (1997)). Let $A = \mathbb{Z}$, and let $P = (P(a, b))_{a, b \in A}$ be a row stochastic $A \times A$ -matrix. Suppose that P is irreducible, i.e., for each pair $a, b \in A$, there is $n > 0$ with $P^n(a, b) > 0$. Suppose that there is a strictly positive function $\pi : A \rightarrow \mathbb{R}_+$ such that $\sum_{a \in A} \pi(a) = \infty$ and $\sum_{a \in A} \pi(a) P(a, b) = \pi(b)$ for all $b \in A$. Let T denote the restriction of the left shift to X_P . Define a measure μ on X_P by setting

$$\mu([a]_i^j) = \pi(a_i) P(a_i, a_{i+1}) \cdots P(a_{j-1}, a_j).$$

Then μ is infinite and σ -finite and T preserves μ . By Kakutani and Parry (1963), if $\sum_{n=1}^{\infty} P^n(0, 0) = \infty$, then T is ergodic. We call the system (X_P, μ, T) the *infinite Markov shift associated with (P, π)* .

Nonsingular Poisson Suspensions

Let $(X, \mathcal{B}, \mu)$ be a σ -finite Lebesgue space with an infinite nonatomic measure. Let $\mathcal{B}_0 \subset \mathcal{B}$ be the collection of subsets of finite measure. Denote by X^* the space of measures ω of the form $\omega = \sum_{i \in I} \delta_{x_i}$, where I is a countable set. Endow X^* with the smallest σ -algebra $\mathcal{B}^*$ such that the maps

$$N_A : X^* \ni \omega \mapsto \omega(A) \in \mathbb{Z}_+ \sqcup \{\infty\}$$

are all measurable, $A \in \mathcal{B}_0$. Let μ^* be the (only) probability measure on $\mathcal{B}^*$ such that

- $\mu^* \circ N_A^{-1}$ has the Poisson distribution with parameter $\mu(A)$, $A \in \mathcal{B}_0$.
- If $A, B \in \mathcal{B}_0$ and $A \cap B = \emptyset$, then the maps N_A and N_B are independent.

Let T be a nonsingular invertible transformation of $(X, \mathcal{B}, \mu)$. Define a transformation $T_*: X^* \rightarrow X^*$ by setting: $T_*\omega := \omega \circ T^{-1}$. If T_* is μ^* -nonsingular, then $(X^*, \mathcal{B}^*, \mu^*, T_*)$ is called *the nonsingular Poisson suspension* of $(X, \mathcal{B}, \mu, T)$ (Danilenko et al. 2022a).

Theorem 3.1 (Danilenko et al. 2022a). T_* is μ^* -nonsingular if and only if $\sqrt{\frac{d\mu \circ T}{d\mu}} - 1 \in L^2(\mu)$. If $\frac{d\mu \circ T}{d\mu} - 1 \in L^1(\mu)$ then T_* is μ^* -nonsingular and

$$\frac{d\mu^* \circ T_*}{d\mu^*}(\omega) = e^{-\int_X \left(\frac{d\mu \circ T}{d\mu} - 1\right) d\mu} \prod_{\omega(\{x\})=1} \frac{d\mu \circ T}{d\mu}(x)$$

at a.e. $\omega \in X^*$.

Nonsingular Gaussian Transformations

Let γ stand for the normalized Gaussian measure on $\mathbb{R}$: $d\gamma(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$. Let $(X, \mu) := (\mathbb{R}, \gamma)^{\mathbb{N}}$. Denote by O the group of orthogonal operators in a separable infinite dimensional real Hilbert space $\mathcal{H}$. Fix an orthonormal base in $\mathcal{H}$. Then we can identify $\mathcal{H}$ with $\ell^2(\mathbb{N})$ and every operator from O with an infinite $\mathbb{N} \times \mathbb{N}$ -matrix. Every $O \in O$ determines a Borel transformation $T_O: X \rightarrow X$ by the formula

$$T_O x := ((T_O x)_n)_{n \in \mathbb{N}}, \text{ where } (T_O x)_n := \sum_{m=1}^{\infty} O_{n,m} x_m, n \in \mathbb{N}.$$

More precisely, T_O is defined μ -almost everywhere and it is invertible (mod 0). Moreover, T_O preserves μ . The set $\{T_O | O \in O\}$ is exactly the family of classical (probability preserving) Gaussian transformations, which are well studied in ergodic theory. A wider class of nonsingular Gaussian transformations is defined by Arano, Isono, and Marrakchi in Arano et al. (2021) (see also Danilenko and Lemańczyk 2022 for a different but equivalent presentation of their concepts). For each $y \in X$, consider a transformation $S_y: X \ni x \mapsto x + y \in X$. By the Cameron-Martin theorem, S_y is μ -nonsingular if and only if $y \in \ell^2(\mathbb{N})$ and

$$\frac{d\mu \circ S_y^{-1}}{d\mu}(x) = e^{-\frac{1}{2}\|y\|_2^2 + \sum_{n=1}^{\infty} x_n y_n} \text{ at}$$

$$\mu - \text{a.e. } x = (x_n)_{n=1}^{\infty} \in X \text{ and } y = (y_n)_{n=1}^{\infty} \in \ell^2.$$

It is easy to see that S_y is totally dissipative. It is straightforward to verify that $T_O S_y T_O^{-1} = S_{Oy}$. Denote the affine group $\mathcal{H} \rtimes O$ of $\mathcal{H}$ by $\text{Aff } \mathcal{H}$. Then for each $(h, O) \in \text{Aff } \mathcal{H}$, consider a transformation $G_{h,O} := S_h T_O$ of (X, μ) . It is called *a nonsingular Gaussian transformation*. Thus, each nonsingular Gaussian transformation is a composition of a classical probability preserving Gaussian transformation and a nonsingular totally dissipative “rotation.” It is obvious that

$$\frac{d\mu \circ G_{h,O}^{-1}}{d\mu}(x) = \frac{d\mu \circ S_h^{-1}}{d\mu}(x) = e^{-\|h\|_2^2 + \sum_{n=1}^{\infty} x_n h_n} \text{ at}$$

$$\mu - \text{a.e. } x = (x_n)_{n=1}^{\infty} \in X.$$

IDPFT Transformations

Definition 3.2 (Danilenko and Lemańczyk 2019). Let T_n be an ergodic nonsingular invertible transformation of a standard probability space $(X_n, \mathfrak{B}_n, \mu_n)$ for each $n \in \mathbb{N}$. Denote by T the infinite direct product of T_n , $n \in \mathbb{N}$, acting on the infinite product space $(X, \mathfrak{B}, \mu) := \otimes_{n \in \mathbb{N}} (X_n, \mathfrak{B}_n, \mu_n)$. If T is μ -nonsingular and each T_n is of finite type, i.e., there exists a μ_n -equivalent probability measure ν_n which is invariant under T_n for each $n \in \mathbb{N}$, then T is called an *infinite direct product of finite types (IDPFT)*.

Let $\varphi_n := \frac{d\mu_n}{d\nu_n}$. It follows from the Kakutani criterion (Kakutani 1948) that T is μ -nonsingular if and only if $\prod_{n=1}^{\infty} \int_{X_n} \sqrt{\varphi_n} \cdot \varphi_n \circ T_n > 0$. Moreover, μ is mutually singular with the T -invariant probability $\nu := \otimes_{n \in \mathbb{N}} \nu_n$ if and only if $\prod_{n=1}^{\infty} \int_{X_n} \sqrt{\varphi_n} d\nu_n = 0$. If T is μ -nonsingular, then

$$\frac{d\mu \circ T}{d\mu}(x) = \prod_{n=1}^{\infty} \frac{d\mu_n \circ T_n}{d\mu_n}(x_n) \text{ at}$$

$$\text{a.e. } x = (x_n)_{n=1}^{\infty} \in X.$$

Natural Extensions of Nonsingular Endomorphisms

Let $(X, \mathcal{B}, \mu)$ be a σ -finite standard measure space. A *nonsingular endomorphism* is a

measurable map $R : X \rightarrow X$ such that $\mu(A) = 0$ if and only if $\mu(R^{-1}A) = 0$. Suppose that μ is σ -finite on $R^{-1}\mathcal{B}$. We define the *Radon-Nikodym derivative* ω_1^μ of R by setting $\omega_1^\mu = \frac{d\mu}{d\mu \circ R^{-1}} \circ R$. It was shown in (Silva 1988; Silva and Thieullen 1995) that there exists a σ -finite standard measure space $(X^*, \mathcal{B}^*, \mu^*)$, an invertible μ^* -nonsingular transformation R^* , and a Borel map $\pi : X^* \rightarrow X$ such that the following hold: $\mu^* \circ \pi^{-1} = \mu$, $\pi R^* = R\pi$, $\omega_1^{\mu^*}$ is $\pi^{-1}(\mathcal{B})$ -measurable and $\bigvee_{n \geq 0} R^n \pi^{-1}(\mathcal{B}) = \mathcal{B}^*$. The dynamical system $(X^*, \mathcal{B}^*, R^*, \mu^*)$ is defined uniquely (up to a natural isomorphism) and called *the natural extension* of R . It coincides with the standard Rokhlin definition of the natural extension in the case where R preserves μ and μ is finite.

Theorem 3.3 R^* is conservative if and only if R is μ -recurrent, i.e.,

$$\sum_{i \geq 0} h \circ R^i \omega_i = +\infty \quad \text{a.e., where } \omega_i = \prod_{j=0}^{i-1} \omega_1^{\mu^*} \circ R^j$$

for each integrable function $h > 0$. Moreover, if R is μ -recurrent, then R^* is ergodic if and only if R is ergodic (Silva 1988; Silva and Thieullen 1995).

Let R be a nonsingular one-sided Bernoulli shift $(X, \mu) = \bigotimes_{n=1}^{\infty} (A, \mu_n)$. Then the natural extension of R is isomorphic to the two-sided nonsingular Bernoulli shift T on $(X^*, \mu^*) = \bigotimes_{n=-\infty}^{\infty} (A, \mu_n^*)$, where $\mu_n^* = \mu_n$ if $n > 0$ and $\mu_n^* = \mu_1$ if $n \leq 0$. The corresponding projection $\pi : X^* \rightarrow X$ is the natural projection, i.e., $\pi(\dots, a_{-1}, a_0, a_1, a_2, \dots) := (a_1, a_2, \dots)$.

Topological Groups $\text{AUT}(X, \mu)$, $\text{AUT}_2(X, \mu)$ and $\text{AUT}_1(X, \mu)$

4.1. Let $(X, \mathcal{B}, \mu)$ be a standard probability space, and let $\text{Aut}(X, \mu)$ denote the group of all nonsingular transformations of X . Let ν be a finite or σ -finite measure equivalent to μ ; the subgroup of the ν -preserving transformations is denoted by $\text{Aut}_0(X, \nu)$. Then $\text{Aut}(X, \mu)$ is a simple group (Eigen 1981), and it has no outer automorphisms

(Eigen 1982). Ryzhikov showed (Ryzhikov 1993) that every element of this group is a product of three involutions (i.e., transformations of order 2). Moreover, a nonsingular transformation is a product of two involutions if and only if it is conjugate to its inverse by an involution.

Inspired by (Halmos 1956), Ionescu Tulcea (Ionescu Tulcea 1965) and Chacon and Friedman (Chacon and Friedman 1965) introduced the *weak* and the *uniform* topologies, respectively, on $\text{Aut}(X, \mu)$. The weak one – we denote it by d_w – is induced from the weak operator topology on the group of unitary operators in $L^2(X, \mu)$ by the embedding $T \mapsto U_T$ (see §2.3). Then $(\text{Aut}(X, \mu), d_w)$ is a Polish topological group and $\text{Aut}_0(X, \nu)$ is a closed subgroup of $\text{Aut}(X, \mu)$. This topology will not be affected if we replace μ with any equivalent measure. We note that T_n weakly converges to T if and only if $\mu(T_n^{-1} A \Delta T^{-1} A) \rightarrow 0$ for each $A \in \mathcal{B}$ and $d(\mu \circ T_n)/d\mu \rightarrow d(\mu \circ T)/d\mu$ in $L^1(X, \mu)$. For each $p \geq 1$, one can also embed $\text{Aut}(X, \mu)$ into the isometry group of $L^p(X, \mu)$ via a formula similar to (3) but with another power of the Radon-Nikodym derivative in it. The strong operator topology on the isometry group induces the very same weak topology on $\text{Aut}(X, \mu)$ for all $p \geq 1$ (Choksi and Kakutani 1979). Danilenko showed in Danilenko (1995) that $(\text{Aut}(X, \mu), d_w)$ is contractible. It follows easily from the Rokhlin lemma that periodic transformations are dense in $\text{Aut}(X, \mu)$.

It is natural to ask which properties of nonsingular transformations are typical in the sense of Baire category. The following technical lemma (see Friedman 1970; Choksi and Kakutani 1979) is an indispensable tool when considering such problems.

Lemma 4.1 *The conjugacy class of each aperiodic transformation T is dense in $\text{Aut}(X, \mu)$ endowed with the weak topology.*

It follows that $\text{Aut}(X, \mu)$ has the *Rokhlin property*, i.e., there is an element of this group whose conjugacy class is dense in the group. Using Lemma 4.1 and the Hurewicz ergodic theorem, Choksi and Kakutani (Choksi and Kakutani 1979) proved that the ergodic transformations form a dense G_δ in $\text{Aut}(X, \mu)$. The same holds for the

subgroup $\text{Aut}_0(X, \nu)$ (Sachdeva 1971; Choksi and Kakutani 1979). Combined with (Ionescu Tulcea 1965), the above implies that the ergodic transformation of type III is a dense G_δ in $\text{Aut}(X, \mu)$. For further refinement of this statement, we refer to section “Orbit Theory”.

Since the map $T \mapsto T \times \cdots \times T$ (p times) from $\text{Aut}(X, \mu)$ to $\text{Aut}(X^p, \mu^{\otimes p})$ is continuous for each $p > 0$, we deduce that the set $\mathcal{E}_\infty$ of transformations with infinite ergodic index (which means that $T \times \cdots \times T$ (p times) is ergodic for each $p > 0$) is a G_δ in $\text{Aut}(X, \mu)$. It is nonempty by (Kakutani and Parry 1963). Since this $\mathcal{E}_\infty$ is invariant under conjugacy, it is dense in $\text{Aut}(X, \mu)$ by Lemma 4.1. Thus, we obtain that $\mathcal{E}_\infty$ is a dense G_δ . In a similar way, one can show that $\mathcal{E}_\infty \cap \text{Aut}_0(X, \nu)$ is a dense G_δ in $\text{Aut}_0(X, \nu)$ (see also (Sachdeva 1971; Choksi and Kakutani 1979; Choksi and Nadkarni 2000) for original proofs of these claims).

A nonsingular transformation T is called *rigid* if $T^{n_k} \rightarrow \text{Id}$ weakly for some sequence $n_k \rightarrow \infty$. The rigid transformations form a dense G_δ in $\text{Aut}(X, \mu)$. It follows that the set of multiply recurrent nonsingular transformations is residual (Ageev and Silva 2001). A finer result was established in Danilenko and Silva (2004): The set of polynomially recurrent transformations in $\text{Aut}_0(X, \nu)$ is residual in $\text{Aut}_0(X, \nu)$. For the definition of multiple and polynomial recurrence, we refer to §6.5 below.

Given $T \in \text{Aut}(X, \mu)$, we denote the *centralizer* $\{S \in \text{Aut}(X, \mu) \mid ST = TS\}$ of T by $C(T)$. Of

course, $C(T)$ is a closed subgroup of $\text{Aut}(X, \mu)$ and $C(T) \supset \{T^n \mid n \in \mathbb{Z}\}$. In a similar way, if $T \in \text{Aut}_0(X, \nu)$, the *measure-preserving centralizer* $C_0(T) := \text{Aut}_0(X, \nu) \cap C(T)$ of T is a weakly closed subgroup of $\text{Aut}_0(X, \nu)$. The following problems solved (by several authors) for probability-preserving systems are still open for the nonsingular case. The properties are:

- (i) T has square root.
- (ii) T embeds into a flow.
- (iii) T has nontrivial invariant sub- σ -algebra.
- (iv) $C(T)$ contains a torus of arbitrary dimension.

typical (residual) in $\text{Aut}(X, \mu)$ or $\text{Aut}_0(X, \nu)$?

The *uniform* topology on $\text{Aut}(X, \mu)$, finer than d_u , is defined by the metric

$$d_u(T, S) = \mu(\{x : Tx \neq Sx\}) + \mu(\{x : T^{-1}x \neq S^{-1}x\}).$$

This topology is also complete metric. It depends only on the measure class of μ . However, the uniform topology is not separable, and that is why it is of less importance in ergodic theory. We refer to (Chacon and Friedman 1965; Friedman 1970; Choksi and Kakutani 1979; Choksi and Prasad 1983) for the properties of d_u .

4.2. Suppose now that $\mu(X) = \infty$ but μ is σ -finite. We now let

$$\begin{aligned} \text{Aut}_2(X, \mu) &:= \left\{ T \in \text{Aut}(X, \mu) \mid \sqrt{\frac{d\mu \circ T}{d\mu}} - 1 \in L^2(\mu) \right\} \text{ and} \\ \text{Aut}_1(X, \mu) &:= \left\{ T \in \text{Aut}(X, \mu) \mid \frac{d\mu \circ T}{d\mu} - 1 \in L^1(\mu) \right\}. \end{aligned}$$

These groups appear naturally in the study of nonsingular Poisson suspensions (see (3.7) and §9). Define a topology d_2 on $\text{Aut}_2(X, \mu)$ by setting that a sequence $(T_n)_{n=1}^\infty$ converges to T in d_2 if $T_n \rightarrow T$ weakly and $\left\| \sqrt{\frac{d\mu \circ T_n}{d\mu}} - \sqrt{\frac{d\mu \circ T}{d\mu}} \right\|_2 \rightarrow 0$ as $n \rightarrow \infty$. In a similar way, one can define a topology d_1 on $\text{Aut}_1(X, \mu)$: A sequence $(T_n)_{n=1}^\infty$

converges to T in d_1 if $T_n \rightarrow T$ weakly and $\left\| \frac{d\mu \circ T_n}{d\mu} - \frac{d\mu \circ T}{d\mu} \right\|_1 \rightarrow 0$ as $n \rightarrow \infty$.

Theorem 4.2 (Danilenko et al. 2022a)

- $(\text{Aut}_2(X, \mu), d_2)$ is a Polish group.
- $(\text{Aut}_1(X, \mu), d_1)$ is a Polish group.

- The mapping $\chi : \text{Aut}_1(X, \mu) \ni T \mapsto \int_X \left(\frac{d\mu \circ T}{d\mu} - 1 \right) d\mu \in \mathbb{R}$ is a continuous onto homomorphism.
- The short exact sequence

$$\{1\} \rightarrow \ker \chi \rightarrow \text{Aut}_1(X, \mu) \xrightarrow{\chi} \mathbb{R} \rightarrow \{1\}$$

splits.

- $\text{Aut}_1(X, \mu)$ is a dense and meager subgroup of $(\text{Aut}_2(X, \mu), d_2)$.
- χ is not d_2 -continuous.
- The group $\{T_* \mid T \in \text{Aut}_2(X, \mu)\}$ is weakly closed in $\text{Aut}(X^*, \mu^*)$.
- The groups $(\text{Aut}_2(X, \mu), d_2)$ and $(\ker \chi, d_1)$ have the Rokhlin property.
- The group $(\text{Aut}_1(X, \mu), d_1)$ does not have the Rokhlin property.

Orbit Theory

Orbit theory is, in a sense, the most complete part of nonsingular ergodic theory. We present here the seminal Krieger's theorem on orbit classification of ergodic nonsingular transformations in terms of ratio sets and associated flows. Examples of transformations of various types III_λ , $0 \leq \lambda \leq 1$ are given. "Almost continuous" refinement of the orbit equivalence is also considered here. Next, we consider the outer conjugacy problem for automorphisms of the orbit equivalence relations. This problem is solved in terms of a simple complete system of invariants. We discuss also a general theory of cocycles (of nonsingular systems) taking values in locally compact Polish groups and present an important orbit classification theorem for cocycles. This theorem is an analogue of the aforementioned result of Krieger. We complete the section by considering ITPFI-systems and their relation to AT-flows.

Full Groups. Ratio Set and Types III_λ , $0 \leq \lambda \leq 1$

Let T be a nonsingular transformation of a standard probability space $(X, \mathcal{B}, \mu)$. Denote by $\text{Orb}_T(x)$ the T -orbit of x , i.e., $\text{Orb}_T(x) = \{T^n x \mid n \in \mathbb{Z}\}$. The *full group* $[T]$ of T consists of all

transformations $S \in \text{Aut}(X, \mu)$ such that $Sx \in \text{Orb}_T(x)$ for a.a. x . If T is ergodic, then $[T]$ is topologically simple (or even algebraically simple if T is not of type II_∞) (Eigen 1981). It is easy to see that $[T]$ endowed with the uniform topology d_u is a Polish group. If T is ergodic, then $([T], d_u)$ is contractible (Danilenko 1995).

The *ratio set* $r(T)$ of T was defined by Krieger [Kr70], and as we shall see below it is the key concept in the orbit classification (see Definition 2.13). The ratio set is a subset of $[0, +\infty)$ defined as follows: $t \in r(T)$ if and only if for every $A \in \mathcal{B}$ of positive measure and each $\epsilon > 0$ there is a subset $B \subset A$ of positive measure and an integer $k \neq 0$ such that $T^k B \subset A$ and $|\omega_k^\mu(x) - t| < \epsilon$ for all $x \in B$. It is easy to verify that $r(T)$ depends only on the equivalence class of μ and not on μ itself. A basic fact is that $1 \in r(T)$ if and only if T is conservative. Assume now T to be conservative and ergodic. Then $r(T) \cap (0, +\infty)$ is a closed subgroup of the multiplicative group $(0, +\infty)$. Hence $r(T)$ is one of the following sets:

- $\{1\}$.
- $\{0, 1\}$; in this case, we say that T is of type III_0 .
- $\{\lambda^n \mid n \in \mathbb{Z}\} \cup \{0\}$ for $0 < \lambda < 1$; then we say that T is of type III_λ .
- $[0, +\infty)$; then we say that T is of type III_1 .

Krieger showed that $r(T) = \{1\}$ if and only if T is of type II . Hence we obtain a further subdivision of type III into subtypes III_0 , III_λ , $0 < \lambda < 1$, and III_1 . The subset of transformations of type III_1 is a dense $G\delta$ in $\text{Aut}(X, \mu)$ (Choksi et al 1987; Parthasarathy and Schmidt 1977).

Example 5.1

- Fix $\lambda \in (0, 1)$. Let $v_n(0) := 1/(1 + \lambda)$ and $v_n(1) := \lambda/(1 + \lambda)$ for all $n = 1, 2, \dots$. Let T be the nonsingular product odometer associated with the sequence $(2, v_n)_{n=1}^\infty$ (see §3.1). We claim that T is of type III_λ . Indeed, the group Σ of finite permutations of $\mathbb{N}$ acts on X by $(\sigma x)_n = x_{\sigma^{-1}(n)}$ for all $n \in \mathbb{N}$, $\sigma \in \Sigma$, and $x = (x_n)_{n=1}^\infty \in X$. This action preserves μ . Moreover, it is ergodic by the Hewitt-Savage

λ on the cylinder $[0]$ which is of positive measure.

- (ii) Fix positive reals ρ_1 and ρ_2 such that $\log \rho_1$ and $\log \rho_2$ are rationally independent. Let $v_n(0) := 1/(1 + \rho_1 + \rho_2)$, $v_n(1) := \rho_1/(1 + \rho_1 + \rho_2)$, and $v_n(2) := \rho_2/(1 + \rho_1 + \rho_2)$ for all $n = 1, 2, \dots$. Then the nonsingular product odometer associated with the sequence $(3, v_n)_{n=1}^\infty$ is of type III_1 . This can be shown in a similar way as (i).
- (iii) Partition $\mathbb{N}$ into two infinite subsets A and B . Fix a sequence $(\epsilon_n)_{n \in B}$ of positive reals such that $\epsilon_n < 0.4$ for all $n \in B$ and $\sum_{n \in B} \epsilon_n < \infty$. We now let $v_n(0) := 0.5$ if $n \in A$ and $\mu_n(0) := 1 - \epsilon_n$ if $n \in B$. Then the nonsingular product odometer associated with the sequence $(2, v_n)_{n=1}^\infty$ is of type II_∞ . This follows from (Moore 1967).

Nonsingular product odometer of type III_0 will be constructed in Example 5.3 below.

Maharam Extension, Associated Flow, and Orbit Classification of Type III Systems

On $X \times \mathbb{R}$ with the σ -finite measure $\mu \times \kappa$, where $d\kappa(y) = \exp(y)dy$, consider the transformation

$$\tilde{T}(x, y) := \left(Tx, y - \log \frac{d\mu \circ T}{d\mu}(x) \right).$$

We call it the *Maharam extension* of T (see (Maharam 1964), where these transformations were introduced). It is measure-preserving, and it commutes with the flow $S_t(x, y) := (x, y + t)$, $t \in \mathbb{R}$. It is conservative if and only if T is conservative (Maharam 1964). However $\tilde{T}$ is not necessarily ergodic when T is ergodic. Let (Z, ν) denote the space of $\tilde{T}$ -ergodic components. Then $(S_t)_{t \in \mathbb{R}}$ acts nonsingularly on this space. The restriction of $(S_t)_{t \in \mathbb{R}}$ to (Z, ν) is called the *associated flow* of T . The associated flow is ergodic if and only if T is ergodic. It is easy to verify that the isomorphism class of the associated flow is an invariant of the orbit equivalence of the underlying system.

Proposition 5.2

- (i) T is of type II if and only if its associated flow is the translation on $\mathbb{R}$, i.e., $x \mapsto x + t$, $x, t \in \mathbb{R}$.
- (ii) T is of type III_λ , $0 \leq \lambda < 1$ if and only if its associated flow is the periodic flow on the interval $[0, -\log \lambda)$, i.e., $x \mapsto x + t \bmod (-\log \lambda)$.
- (iii) T is of type III_1 if and only if its associated flow is the trivial flow on a singleton or, equivalently, $\tilde{T}$ is ergodic.
- (iv) T is of type III_0 if and only if its associated flow is nontransitive (Hamachi and Osikawa 1981).

Example 5.3 Let $A_n = \{0, 1, \dots, 2^n\}$ and $v_n(0) = 0.5$ and $v_n(i) = 0.5 \cdot 2^{-2^n}$ for all $0 < i \leq 2^n$. Let T be the nonsingular product odometer associated with $(2^{2^n} + 1, v_n)_{n=0}^\infty$. It is straightforward that the associated flow of T is the flow built under the constant function 1 with the probability preserving 2-adic product odometer (associated with $(2, \kappa_n)_{n=1}^\infty$, $\kappa_n(0) = \kappa_n(1) = 0.5$) as the base transformation. In particular, T is of type III_0 .

A natural problem arises: to compute Krieger's type (or the ratio set) for the nonsingular product odometers – the simplest class of nonsingular systems. Some partial progress was achieved in (Araki and Woods 1968; Dooley et al. 1998; Moore 1967; Osikawa 1988), etc. However, in the general setting this problem remains open.

The map $\Psi : \text{Aut}(X, \mu) \ni T \mapsto \tilde{T} \in \text{Aut}(X \times \mathbb{R}, \mu \times \kappa)$ is a continuous group homomorphism. Since the set $\mathcal{E}$ of ergodic transformations on $X \times \mathbb{R}$ is a G_δ in $\text{Aut}(X \times \mathbb{R}, \mu \times \kappa)$ (See § 4), the subset $\Psi^{-1}(\mathcal{E})$ of type III_1 ergodic transformations on X is also G_δ . The latter subset is nonempty in view of Example 5.1(ii). Since it is invariant under conjugacy, we deduce from Lemma 4.1 that the set of ergodic transformations of type III_1 is a dense G_δ in $(\text{Aut}(X, \mu), d_w)$ (Choksi et al. 1987; Parthasarathy and Schmidt 1977).

Now we state the main result of this section – Krieger’s theorem on orbit classification for ergodic transformations of type III . It is a far-reaching generalization of the basic result by H. Dye: Any two ergodic probability-preserving transformations are orbit equivalent (Dye 1959).

Theorem 5.4 (Orbit equivalence for type III systems (Krieger 1969, 1970, 1972, 1976a, b)). *Two ergodic transformations of type III are orbit equivalent if and only if their associated flows are isomorphic. In particular, for a fixed $0 < \lambda \leq 1$, any two ergodic transformations of type III_λ are orbit equivalent.*

The original proof of this theorem is rather complicated. Simpler treatment of it can be found in Hamachi and Osikawa (1981) and Katznelson and Weiss (1991).

We also note that every free ergodic flow can be realized as the associated flow of a type III_0 transformation. However it is somewhat easier to construct a $\mathbb{Z}^2$ -action of type III_0 whose associated flow is the given one. For this, we take an ergodic nonsingular transformation Q on a probability space $(Z, \mathcal{B}, \lambda)$ and a measure-preserving transformation R of an infinite σ -finite measure space $(Y, \mathcal{F}, \nu)$ such that there is a continuous homomorphism $\pi: \mathbb{R} \rightarrow C(R)$ with $(d\nu \circ \pi(t)/d\nu)(y) = \exp(t)$ for a.a. y (for instance, take a type III_1 transformation T and put $R := \tilde{T}$ and $\pi(t) := S_t$). Let $\varphi: Z \rightarrow \mathbb{R}$ be a Borel map with $\inf_Z \varphi > 0$. Define two transformations R_0 and Q_0 of $(Z \times Y, \lambda \times \nu)$ by setting:

$$R_0(x, y) := (x, Ry), \quad Q_0(x, y) = (Qx, U_{xy}),$$

where $U_x = \pi(\varphi(x) - \log(d\mu \circ Q/d\mu)(x))$. Notice that R_0 and Q_0 commute. The corresponding $\mathbb{Z}^2$ -action generated by these transformations is ergodic. Take any transformation $V \in \text{Aut}(Z \times Y, \lambda \times \nu)$ whose orbits coincide with the orbits of the $\mathbb{Z}^2$ -action. (According to Connes et al. (1981), any ergodic nonsingular action of any countable amenable group is orbit equivalent to a single transformation.) It is now easy to verify that the associated flow of V is the special flow built under $\varphi \circ Q^{-1}$ with the base transformation Q^{-1} . Then V is of type III_0 . Since Q and φ are arbitrary, we deduce the following from Theorem 2.17.

Theorem 5.5 *Every nontransitive ergodic flow is an associated flow of an ergodic transformation of type III_0 .*

In Krieger (1976b), Krieger introduced a map Φ as follows. Let T be an ergodic transformation of type III_0 . Then the associated flow of T is a flow built under function with a base transformation $\Phi(T)$. We note that the orbit equivalence class of $\Phi(T)$ is well defined by the orbit equivalent class of T . If $\Phi^n(T)$ fails to be of type III_0 for some $1 \leq n < \infty$, then T is said to *belong to Krieger’s hierarchy*. For instance, the transformation constructed in Example 5.3 belongs to Krieger’s hierarchy. Connes gave in Connes (1975) an example of T such that $\Phi(T)$ is orbit equivalent to T (see also Hamachi and Osikawa (1981; Giordano and Skandalis 1985a). Hence T is not in Krieger’s hierarchy.

Almost Continuous Orbit Equivalence

In this subsection, by a dynamical system we mean a quadruple (X, τ, μ, T) , where (X, τ) is a Polish space, μ is a nonatomic Borel measure of full support, and T is a nonsingular ergodic homeomorphism of X such that the function $\omega_1: X \rightarrow \mathbb{R}$ is continuous (has a continuous version).

Definition 5.6 Two dynamical systems (X, τ, μ, T) and (X', τ', μ', T') are *almost continuously orbit equivalent* if there are dense invariant G_δ subsets $X_0 \subset X$ and $X'_0 \subset X'$ of full measure and a homeomorphism $\varphi: X_0 \rightarrow X'_0$ such that

- $\varphi(\{T^n x \mid n \in \mathbb{Z}\}) = \{(T')^n \varphi(x) \mid n \in \mathbb{Z}\}$ at every $x \in X_0$.
- $\mu \circ \varphi^{-1} \sim \mu'$ and the Radon-Nikodym derivative $\frac{d\mu \circ \varphi^{-1}}{d\mu'}$ is (can be chosen) continuous.
- letting $S := \varphi^{-1} T' \varphi$, we have $Tx = S^{n(x)}x$ and $Sx = T^{m(x)}x$, where n and m are continuous on X_0 .

We note that in the case where X and X' are infinite product spaces, T and T' preserve μ and μ' , respectively, and we omit the requirement that X_0 and X'_0 are G_δ then the above definition of φ is equivalent to the “finitary” equivalence from the

celebrated work of Keane and Smorodinsky (1979). It was shown by del Junco and Şahin (2009) that any two ergodic probability-preserving homeomorphisms of Polish spaces are almost continuously orbit equivalent. The same is true for any ergodic homeomorphisms preserving infinite σ -finite local measures (del Junco and Şahin 2009). In Danilenko and del Junco (2011), a topological analogue $r_{\text{top}}(T)$ of $r(T)$ was introduced. It is a closed subgroup of $\mathbb{R}$ which contains $r(T)$, and it is invariant under the almost continuous orbit equivalence. In Danilenko and del Junco (2011), two-type III homeomorphisms were constructed which are measure-theoretically orbit equivalent but not almost continuously orbit equivalent (their r_{top} -invariants are different).

Theorem 5.7 *Let (X, τ, μ, T) and (X', τ', μ', T') be ergodic nonsingular homeomorphisms of Polish spaces. If the two systems are either (Danilenko and del Junco 2011)*

- (i) *Of type III_λ with $0 < \lambda < 1$ and $r_{\text{top}}(T) = r_{\text{top}}(T') = \log \lambda \cdot \mathbb{Z}$ or*
- (ii) *Of type III_1 ,*

then they are almost continuously orbit equivalent.

Characterization of almost continuous orbit equivalence for homeomorphisms of type III_0 remains an open problem.

Normalizer of the Full Group. Outer Conjugacy Problem

Let

$$N[T] = \{R \in \text{Aut}(X, \mu) | R[T]R^{-1} = [T]\},$$

i.e., $N[T]$ is the *normalizer* of the full group $[T]$ in $\text{Aut}(X, \mu)$. We note that a transformation R belongs to $N[T]$ if and only if $R(\text{Orb}_T(x)) = \text{Orb}_T(Rx)$ for a.a. x . To define a topology on $N[T]$, consider the T -orbit equivalence relation $\mathcal{R}_T \subset X \times X$ and a σ -finite measure $\mu_{\mathcal{R}}$ on $\mathcal{R}_T$ given by $\mu_{\mathcal{R}_T} = \int_X \sum_{y \in \text{Orb}_T(x)} \delta_{(x,y)} d\mu(x)$. For $R \in N[T]$, we define a transformation

$i(R) \in \text{Aut}(\mathcal{R}_T, \mu_{\mathcal{R}_T})$ by setting $i(R)(x, y) := (Rx, Ry)$. Then the map $R \mapsto i(R)$ is an embedding of $N[T]$ into $\text{Aut}(\mathcal{R}_T, \mu_{\mathcal{R}_T})$. Denote by τ the topology on $N[T]$ induced by the weak topology on $\text{Aut}(\mathcal{R}_T, \mu_{\mathcal{R}_T})$ via i (Danilenko 1995). Then $(N[T], \tau)$ is a Polish group. A sequence R_n converges to R in $(N[T], \tau)$ if $R_n \rightarrow R$ weakly (in $\text{Aut}(X, \mu)$) and $R_n T R_n^{-1} \rightarrow R T R^{-1}$ uniformly (in $[T]$).

Given $R \in N[T]$, denote by $\tilde{R}$ the Maharam extension of R . Then $\tilde{R} \in N[\tilde{T}]$ and it commutes with $(S_t)_{t \in \mathbb{R}}$. Hence it defines a nonsingular transformation mod R on the space (Z, ν) of the associated flow $W = (W_t)_{t \in \mathbb{R}}$ of T . Moreover, mod R belongs to the centralizer $C(W)$ of W in $\text{Aut}(Z, \nu)$. Note that $C(W)$ is a closed subgroup of $(\text{Aut}(Z, \nu), d_w)$.

Let T be of type II_∞ , and let μ' be the invariant σ -finite measure equivalent to μ . If $R \in N[T]$, then it is easy to see that the Radon-Nikodym derivative $d\mu' \circ R/d\mu'$ is invariant under T . Hence it is constant, say c . Then mod $R = \log c$.

Theorem 5.8 *If T is of type III, then the map mod : $N[T] \rightarrow C(W)$ is a continuous onto homomorphism. The kernel of this homomorphism is the τ -closure of $[T]$. Hence the quotient group $N[T]/\overline{[T]}^\tau$ is (topologically) isomorphic to $C(W)$. In particular, $\overline{[T]}^\tau$ is cocompact in $N[T]$ if and only if W is a finite measure-preserving flow with a pure point spectrum (Hamachi and Osikawa 1981; Hamachi 1981a).*

The following theorem describes the homotopical structure of normalizers.

Theorem 5.9 *Let T be of type II or III, $0 \leq \lambda < 1$. The group $\overline{[T]}^\tau$ is contractible. $N[T]$ is homotopically equivalent to $C(W)$. In particular, $N[T]$ is contractible if T is of type II. If T is of type III_λ with $0 < \lambda < 1$, then $\pi_1(N[T]) = \mathbb{Z}$ (Danilenko 1995).*

The *outer period* $p(R)$ of $R \in N[T]$ is the smallest positive integer n such that $R^n \in [T]$. We write $p(R) = 0$ if no such n exists.

Two transformations R and R' in $N[T]$ are called *outer conjugate* if there are transformations $V \in N[T]$ and $S \in [T]$ such that $V R V^{-1} = R' S$. The following theorem provides convenient (for

verification) necessary and sufficient conditions for the outer conjugacy.

Theorem 5.10 (Connes and Krieger 1977) for type II and (Bezuglyi and Golodets 1985) for type III). Transformations $R, R' \in N[T]$ are outer conjugate if and only if $p(R) = p(R')$ and $\text{mod } R$ is conjugate to $\text{mod } R'$ in the centralizer of the associated flow of T .

We note that in the case T is of type II, the second condition in the theorem is just $\text{mod } R = \text{mod } R'$. It is always satisfied when T is of type II₁.

Cocycles of Dynamical Systems. Weak Equivalence of Cocycles

Let G be a locally compact Polish group and λ_G a left Haar measure on G . A Borel map $\varphi : X \rightarrow G$ is called a *cocycle* of T . Two cocycles φ and φ' are *cohomologous* if there is a Borel map $b : X \rightarrow G$ such that

$$\varphi'(x) = b(Tx)^{-1} \varphi(x) b(x)$$

for a.a. $x \in X$. A cocycle cohomologous to the trivial one is called a *coboundary*. Given a dense subgroup $G' \subset G$, every cocycle is cohomologous to a cocycle with values in G' (Golodets and Sinel'shchikov 1994). Each cocycle φ extends to a (unique) map $\alpha_\varphi : \mathcal{R}_T \rightarrow G$ such that $\alpha_\varphi(Tx, x) = \varphi(x)$ for a.a. x and $\alpha_\varphi(x, y)\alpha_\varphi(y, z) = \alpha_\varphi(x, z)$ for a.a. $(x, y), (y, z) \in \mathcal{R}_T$. α_φ is called the *cocycle of $\mathcal{R}_T$ generated by φ* . Moreover, φ and φ' are cohomologous via b as above if and only if α_φ and $\alpha_{\varphi'}$ are cohomologous via b , i.e., $\alpha_\varphi(x, y) = b(x)^{-1} \alpha_{\varphi'}(x, y) b(y)$ for $\mu_{\mathcal{R}_T}$ -a.a. $(x, y) \in \mathcal{R}_T$. The following notion was introduced by Golodets and Sinel'shchikov (Golodets and Sinel'shchikov 1983, 1994): Two cocycles φ and φ' are *weakly equivalent* if there is a transformation $R \in N[T]$ such that the cocycles α_φ and $\alpha_{\varphi'} \circ (R \times R)$ of $\mathcal{R}_T$ are cohomologous. Let $\mathcal{M}(X, G)$ denote the set of Borel maps from X to G . It is a Polish group when endowed with the topology of convergence in measure. Since T is ergodic, it is easy to deduce from Rokhlin's lemma that the cohomology class of any cocycle is dense in $\mathcal{M}(X, G)$. Given $\varphi \in \mathcal{M}(X, G)$, we define the *φ -skew product*

extension T_φ of T acting on $(X \times G, \mu \times \lambda_G)$ by setting $T_\varphi(x, g) := (Tx, \varphi(x)g)$. Thus, Maharam extension is (isomorphic to) the Radon-Nikodym cocycle-skew product extension. We now specify some basic classes of cocycles (Schmidt 1977a; Bezuglyi and Golodets 1991; Golodets and Sinel'shchikov 1994; Danilenko 1998):

- (i) φ is called *transient* if T_φ is totally dissipative.
- (ii) φ is called *recurrent* if T_φ is conservative (equivalently, φ is not transient).
- (iii) φ has *dense range* in G if T_φ is ergodic.
- (iv) φ is called *regular* if φ cobounds with dense range into a closed subgroup H of G (then H is defined up to conjugacy).

These properties are invariant under the cohomology and the weak equivalence. The Radon-Nikodym cocycle ω_1 is a coboundary if and only if T is of type II. It is regular if and only if T is of type II or III _{λ} , $0 < \lambda \leq 1$. It has dense range (in the multiplicative group $\mathbb{R}_+^*$) if and only if T is of type III₁. Notice that ω_1 is never transient (since T is conservative).

In case G is Abelian, Schmidt introduced in Schmidt (1984) an invariant $R(\varphi) := \{g \in G \mid \varphi - g \text{ is recurrent}\}$. He showed in particular that

- (i) $R(\varphi)$ is a cohomology invariant.
- (ii) $R(\varphi)$ is a Borel set in G .
- (iii) $R(\log \omega_1) = \{0\}$ for each aperiodic conservative T .
- (iv) There are cocycles φ such that $R(\varphi)$ and $G \setminus R(\varphi)$ are dense in G .
- (v) If $\mu(X) = 1$, $\mu \circ T = \mu$ and $\varphi : X \rightarrow \mathbb{R}$ is integrable, then $R(\varphi) = \{\int \varphi d\mu\}$.

We note that (v) follows from Atkinson theorem (Atkinson 1976). A nonsingular version of this theorem was established in Ullman (1987): If T is ergodic and μ -nonsingular and $f \in L^1(\mu)$, then

$$\lim_{n \rightarrow \infty} \inf \left| \sum_{j=0}^{n-1} f(T^j x) \omega_j(x) \right| = 0 \text{ for a.a. } x$$

if and only if $\int f d\mu = 0$.

Since T_φ commutes with the action of G on $X \times G$ by inverted right translations along the second coordinate, this action induces an ergodic G -action $W_\varphi = (W_\varphi(g))_{g \in G}$ on the space (Z, ν) of T_φ -ergodic components. It is called the *Mackey range (or Poincaré flow)* of φ (Mackey 1966; Feldman and Moore 1977; Schmidt 1977a). We note that φ is regular (and cobounds with dense range into $H \subset G$) if and only if W_φ is transitive (and H is the stabilizer of a point $z \in Z$, i.e., $H = \{g \in G \mid W_\varphi(g)z = z\}$). Hence every cocycle taking values in a compact group is regular.

It is often useful to consider the *double cocycle* $\varphi_0 := \varphi \times \omega_1$ instead of φ . It takes values in the group $G \times \mathbb{R}_+^*$. Since T_{φ_0} is exactly the Maharam extension of T_φ it follows from Maharam (1964) that φ_0 is transient or recurrent if and only if φ is transient or recurrent, respectively.

Theorem 5.11 (Orbit classification of cocycles (Golodets and Sinel'shchikov 1994)). Let $\varphi, \varphi' : X \rightarrow G$ be two recurrent cocycles of an ergodic transformation T . They are weakly equivalent if and only if their Mackey ranges W_{φ_0} and $W_{\varphi'_0}$ are isomorphic.

Another proof of this theorem was presented in Fedorov (1985).

Theorem 5.12 Let T be an ergodic nonsingular transformation. Then there is a cocycle of T with dense range in G if and only if G is amenable.

It follows that if G is amenable then the subset of cocycles of T with dense range in G is a dense G_δ in $\mathcal{M}(X, G)$ (just adapt the argument following Example 5.3). The “only if” part of Theorem 5.12 was established in Zimmer (1978). The “if” part was considered by many authors in particular cases: G is compact Zimmer (1977), G is solvable or amenable almost connected Golodets and Sinel'shchikov (1985), etc. The general case was proved in Golodets and Sinel'shchikov (1983) and Herman (1979a) (see also a refinement in (Aaronson and Weiss 2004)).

We note that the “if” part in Theorem 5.12 can be refined in the case where G is a compactly generated Abelian group.

Theorem 5.13 Let T be an ergodic nonsingular transformation. If G is a compactly generated [FIA]-group, then there is a bounded ergodic cocycle φ of T with values in G (Danilenko 2021).

We recall that a locally compact Polish group G is called a [FIA]-group if the group of inner automorphisms of G is relatively compact in the group of all automorphisms of G furnished with the natural topology (Grosser and Moskowitz 1971). Of course, each Abelian group is [FIA]. The cocycle φ is *bounded* if there is a compact subset K in G such that φ takes values in K .

Theorem 5.5 is a particular case of the following result.

Theorem 5.14 Let G be amenable. Let V be an ergodic nonsingular action of $G \times \mathbb{R}_+^*$. Then there is an ergodic nonsingular transformation T and a recurrent cocycle φ of T with values in G such that V is isomorphic to the Mackey range of the double cocycle φ_0 (Golodets and Sinel'shchikov 1990; Fedorov 1985; Adams et al. 1994).

Given a cocycle $\varphi \in \mathcal{M}(X, G)$ of T , we say that a transformation $R \in N[T]$ is *compatible* with φ if the cocycles α_φ and $\alpha_\varphi \circ (R \times R)$ of $\mathcal{R}_T$ are cohomologous. Denote by $D(T, \varphi)$ the group of all such R . It has a natural Polish topology which is stronger than τ (Danilenko and Golodets 1996). Since $[T]$ is a normal subgroup in $D(T, \varphi)$, one can consider the outer conjugacy equivalence relation inside $D(T, \varphi)$. It is called *φ -outer conjugacy*. Suppose that G is Abelian. Then an analogue of Theorem 5.10 for the φ -outer conjugacy is established in Danilenko and Golodets (1996). Also, the cocycles φ with $D(T, \varphi) = N[T]$ are described there.

ITPFI Transformations and AT-Flows

A nonsingular transformation T is called *ITPFI* (This abbreviates “infinite tensor product of factors of type I” (came from the theory of von Neumann algebras).) if it is orbit equivalent to a nonsingular product odometer (associated to a sequence $(m_n, \nu_n)_{n=1}^\infty$, see § 3.1). If the sequence m_n can be chosen bounded, then T is called *ITPFI of bounded type*. If $m_n = 2$ for all n , then T is called *ITPFI₂*. By Giordano and Skandalis (1985b),

every ITPFI-transformation of bounded type is ITPFI₂. In view of Theorem 5.4 and Example 5.1, every ergodic transformation of type II or III_λ with $0 < \lambda \leq 1$ is ITPFI₂.

A remarkable characterization of ITPFI transformations in terms of their associated flows was obtained by Connes and Woods (1985). We first single out a class of ergodic flows. A nonsingular flow $V = (V_t)_{t \in \mathbb{R}}$ on a space (Ω, ν) is called *approximate transitive (AT)* if given $\epsilon > 0$ and $f_1, \dots, f_n \in L^1_+(X, \mu)$, there exists $f \in L^1_+(X, \mu)$ and $\lambda_1, \dots, \lambda_n \in L^1_+(\mathbb{R}, dt)$ such that

$$\left\| f_j - \int_{\mathbb{R}} f \circ V_t \frac{d\nu \circ V_t}{d\nu} \lambda_j(t) dt \right\|_1 < \epsilon$$

for all $1 \leq j \leq n$. A flow built under a constant ceiling function with a funny rank-one (Ferenczi 1985) probability preserving base transformation is AT (Connes and Woods 1985). In particular, each ergodic finite measure-preserving flow with a pure point spectrum is AT.

Theorem 5.15 *An ergodic nonsingular transformation is ITPFI if and only if its associated flow is AT (Connes and Woods 1985).*

The original proof of this theorem was given in the framework of von Neumann algebras theory. A simpler, purely measure-theoretical proof was given later in Hawkins (1990b) (the “only if” part) and Hamachi (1992) (the “if” part). It follows from Theorem 5.15 that every ergodic flow with pure point spectrum is the associated flow of an ITPFI transformation. This was refined recently in Berendschot and Vaes (2022b): Every ergodic flow with a pure point spectrum is the associated flow of an ITPFI₂ transformation. This fact was proved earlier in Hamachi and Osikawa (1986) only for flows whose spectrum is $\theta\Gamma$, where Γ is a subgroup of $\mathbb{Q}$ and $\theta \in \mathbb{R}$. The existence of ITPFI transformations which are not of bounded type was shown in Krieger (1972).

Krieger introduced an invariant for the orbit equivalence, called *property A*, and showed that each product odometer satisfies property A. He also constructed an ergodic nonsingular

transformation which does not satisfy this property (Krieger 1972). Hence this transformation is not ITPFI. Though not every ergodic transformation is orbit equivalent to a nonsingular product odometer, a “weaker” form of this statement holds.

Theorem 5.16 *Each ergodic nonsingular transformation is orbit equivalent to a Markov odometer (see §3.2) (Dooley and Hamachi 2003a).*

In Dooley and Hamachi (2003b), an explicit example of a non-ITPFI ergodic Markov odometer (not satisfying property A) was constructed. Later Munteanu in Munteanu (2012) exhibited an ergodic non-ITPFI transformation satisfying property A. In Joita and Munteanu (2014), it was constructed an explicit example of a non-AT nonsingular flow W built under a function and over a nonsingular product odometer. Hence every nonsingular ergodic transformation whose associated flow is isomorphic to W is non-ITPFI.

Mixing Notions and Multiple Recurrence

The study of mixing and multiple recurrence are central topics in classical ergodic theory (Cornfeld et al. 1982; Furstenberg 1981). Unfortunately, these notions are considerably less “smooth” in the world of nonsingular systems. The very concepts of any kind of mixing and multiple recurrence are not well understood in view of their ambiguity. Below we discuss nonsingular systems possessing a surprising diversity of such properties that seem equivalent but are different indeed.

Weak Mixing

Let T be an ergodic conservative nonsingular transformation. A number $\lambda \in \mathbb{C}$ is an L^∞ -*eigenvalue* for T if there exists a nonzero $f \in L^\infty$ so that $f \circ T = \lambda f$ a.e. It follows that $|\lambda| = 1$ and f has constant modulus, which we assume to be 1. Denote by $e(T)$ the set of all L^∞ -eigenvalues of T . T is said to be *weakly mixing* if $e(T) = \{1\}$. We refer to (Aaronson 1997, Theorem 2.7.1) for proof of the following Keane’s ergodic multiplier

theorem: Given an ergodic probability-preserving transformation S , the product transformation $T \times S$ is ergodic if and only if $\sigma_S(e(T)) = 0$, where σ_S denotes the measure of (reduced) maximal spectral type of the unitary U_S (see (3)). It follows that T is weakly mixing if and only if $T \times S$ is ergodic for every ergodic probability-preserving S . While in the finite measure-preserving case this implies that $T \times T$ is ergodic, it was shown in (Aaronson et al. 1979) that there exists a weakly mixing nonsingular T with $T \times T$ not conservative, hence not ergodic. In (Adams et al. 1997), a weakly mixing T was constructed with $T \times T$ conservative but not ergodic. A nonsingular transformation T is said to be *weakly doubly ergodic* (originally called *doubly ergodic* in (Bowles et al. 2001)) if for all sets of positive measure A and B there exists an integer $n > 0$ such that $\mu(A \cap T^{-n}A) > 0$ and $\mu(A \cap T^{-n}B) > 0$. Furstenberg (Furstenberg 1981) showed that for finite measure-preserving transformations weak double ergodicity is equivalent to weak mixing. In (Adams et al. 1997; Bowles et al. 2001), it is shown that for nonsingular transformations weak mixing does not imply weak double ergodicity and weak double ergodicity does not imply that $T \times T$ is ergodic.

T is said to have *ergodic index* k if the Cartesian product of k copies of T is ergodic but the product of $k + 1$ copies of T is not ergodic. If all finite Cartesian products of T are ergodic, then T is said to have *infinite ergodic index*. In a similar way, one can define the *conservative index* of T . Parry and Kakutani (Kakutani and Parry 1963) constructed for each $k \in \mathbb{N} \cup \{\infty\}$ an infinite Markov shift of ergodic index k . We note that for each infinite Markov shift, the ergodic index coincides with the conservative index. We constructed in (Adams and Silva 2015; Danilenko 2016a) infinite measure-preserving rank-one transformations of an arbitrary ergodic index k and infinite conservative index. A stronger property is *power weak mixing*, which requires that for all nonzero integers $k_1, \dots, k_r$ the product $T^{k_1} \times \dots \times T^{k_r}$ is ergodic (Day et al. 1999). The following examples were constructed in (Adams et al. 2001; Danilenko 2001a, 2004; Day et al. 1999):

- (i) Power weakly mixing rank-one transformations
- (ii) Nonpower weakly mixing rank-one transformations with infinite ergodic index
- (iii) Nonpower weakly mixing rank-one transformations with infinite ergodic index and such that $T^{k_1} \times \dots \times T^{k_r}$ are all conservative, $k_1, \dots, k_r \in \mathbb{Z}$,

of types II_∞ and III (and various subtypes of III , see section “Orbit Theory”). Thus, we have the following scale of properties (equivalent to weak mixing in the probability-preserving case), where every next property is strictly stronger than the previous ones:

T is weakly mixing
 $\Leftrightarrow T$ is weakly doubly ergodic
 $\Leftrightarrow T \times T$ is ergodic
 $\Leftrightarrow T \times T \times T$ is ergodic
 $\Leftrightarrow \dots \Leftrightarrow T$ has infinite ergodic index
 $\Leftrightarrow T$ is power weakly mixing.

There is a rank-one weakly doubly ergodic T such that $T \times T$ is nonconservative (Bowles et al. 2001), and there is a rank-one weakly doubly ergodic T such that $T \times T$ is conservative but not ergodic (Loh and Silva 2017). We also mention an example of a power weakly mixing transformation of type II_∞ which embeds into a rank-one flow (Danilenko and Solomko 2009). This result was sharpened in (Danilenko and Park 2011): There is an infinite measure-preserving rank-one flow $(R_t)_{t \in \mathbb{R}}$ such that for each $t \neq 0$, the transformation T_t has infinite ergodic index. Several of these notions have been studied in the context of nonsingular actions of locally compact groups by Glasner and Weiss (2016); we mention one condition that has not yet been discussed though only in the context of transformations. A nonsingular transformation T on a probability space is said to be *ergodic with isometric coefficients* if every factor map onto an isometry of a (separable) metric space is constant a.e. Glasner and Weiss show that if $T \times T$ is ergodic (i.e., T is doubly ergodic), then T is ergodic with isometric coefficients, and that if T is ergodic with isometric coefficients,

then it is weakly mixing. In (Loh and Silva 2017), it is shown that if T is weakly doubly ergodic, then it is ergodic with isometric coefficients. In (Glasner and Weiss 2016), there is an example of a T that is ergodic with isometric coefficients but $T \times T$ is not conservative, hence not ergodic. In (Haddock et al. 2022), it is shown that isometric coefficients do not imply weak doubly ergodic. Further conditions related to weak mixing (in the case of infinite measure-preserving transformations) are discussed in the survey (Adams and Silva 2018).

Rational Ergodicity and Rational Weak Mixing

Let T be a conservative, ergodic measure-preserving transformation of a σ -finite measure space $(X, \mathcal{B}, \mu)$. For a function $f: X \rightarrow X$, let $S_n(f) = \sum_{k=0}^{n-1} f \circ T^k$. T be called *rationally ergodic* (Aaronson 1977a) if there is a subset $F \in \mathcal{B}$, $0 < \mu(F) < \infty$, satisfying a *Renyi inequality*, i.e., there exists a constant $M > 0$ such that for all $n \geq 1$,

$$\int_F (S_n(\mathbb{I}_F))^2 d\mu \leq M \left(\int_F S_n(\mathbb{I}_F) d\mu \right)^2.$$

For a related concept of p -rational ergodicity (i.e., the rational ergodicity in L^p), see Aaronson and Weiss (2018). We now set

$$u_k(F) := \frac{\mu(F \cap T^{-k}F)}{\mu(F)^2} \text{ and } a_n(F) := \sum_{k=0}^{n-1} u_k(F).$$

Theorem 6.1 *If T is rationally ergodic and F satisfies the Renyi inequality, then for all measurable sets A and B contained in F (Aaronson 1977a),*

$$\lim_{n \rightarrow \infty} \frac{1}{a_n(F)} \sum_{k=0}^{n-1} \mu(A \cap T^{-k}B) = \mu(A)\mu(B).$$

An ergodic conservative transformation T is *rationally weakly mixing* (Aaronson 2013) if there exists a measurable set F of positive finite measure such that for all measurable sets A and B contained in F we have

$$\lim_{n \rightarrow \infty} \frac{1}{a_n(F)} \sum_{k=0}^{n-1} |\mu(A \cap T^{-k}B) - \mu(A)\mu(B)u_k(F)| = 0, \quad (5)$$

where $u_k(F)$ and $a_n(F)$ are defined as above. When $\mu(X) = 1$ and we let $F = X$, then $a_n(F) = n$ and (5) becomes the condition equivalent to the weak mixing property for a finite measure-preserving transformation. In infinite measure, however, the rational weak mixing condition is not equivalent to weak mixing as we shall see. If in (5) we drop the absolute values, then this condition defines the notion of *weak rational ergodicity* (Aaronson 1979). Then Theorem 6.1 claims that rational ergodicity implies weak rational ergodicity. If in (5) the sequence (n) is replaced by a subsequence (n_i) , we say T is *subsequence rational weak mixing*. *Subsequence weak rationally ergodic* is defined in a similar way. The transformation T is *boundedly rationally ergodic* (Aaronson 1979) if

$$\sup_{n \geq 1} \left\| \frac{a_n(F)}{S_n(\mathbb{I}_F)} \right\|_{\infty} < \infty,$$

The notions of *subsequence boundedly rationally ergodic* and *subsequence rationally ergodic* are defined when the sequence (n) is replaced by a subsequence (n_i) . It can be seen from the definition that bounded rational ergodicity implies rational ergodicity (and similarly for the subsequence versions). Aaronson (Aaronson 1979) showed that rational ergodicity does not imply bounded rational ergodicity, and more recently it was shown by Adams and Silva (Adams and Silva 2016) that weak rational ergodicity does not imply rational ergodicity.

The following theorem was proved in (Aaronson 1979) for the weakly rationally ergodic transformations. The same argument works in the more general case.

Theorem 6.2 *Each subsequence weakly rationally ergodic transformation T of (X, μ) is non-squashable, i.e., each nonsingular transformation commuting with T preserves μ .*

There are several examples of rationally ergodic transformations which are infinite Markov

shifts; see Aaronson (1997). More recently, it was shown in Dai et al. (2015) and Bozgan et al. (2015) that rank-one (infinite measure-preserving) transformations are subsequence boundedly rationally ergodic. The first version of Bozgan et al. (2015) has a proof that the rank-one transformations are subsequence weakly rationally ergodic; a simpler proof was found in Danilenko (2016b), where this property is also established for the class of funny rank one transformations and the class of ergodic transformations of *balanced finite rank*. (A transformation is called of balanced finite rank if it is of finite rank and the bases of the Rokhlin towers on the n -th step of the cutting-and-stacking inductive construction have asymptotically comparable measures as $n \rightarrow \infty$.) Therefore all these transformations are nonsquashable in view of Theorem 6.2.

The rank-one transformations for which the sequence of cuts $(r_n)_{n=1}^\infty$ is bounded are boundedly rationally ergodic (Aaronson et al. 2017; Dai et al. 2015; Bozgan et al. 2015). As for the examples of rationally weakly mixing transformations, Aaronson (2013) shows that Markov shifts with certain conditions on their associated renewal sequences are rationally weakly mixing, and Dai et al. (2015) give rank-one examples. Subsequence rational weak mixing and rational weak mixing for products of powers have been studied in Aaronson (2013) and Adams (2015).

We have the following implications for rational weak mixing.

Theorem 6.3 *If a transformation is sequentially rationally weakly mixing, then it is weakly mixing (Aaronson 2013).*

Theorem 6.4 *If a transformation is rationally weakly mixing, then it is weakly doubly ergodic (Bozgan et al. 2015).*

It is an open problem whether weak double ergodicity implies rational weak mixing. Aaronson (2013) asked if weak rational ergodicity and weak mixing imply rational weak mixing. This was answered in negative in (Dai et al. 2015), where was constructed an example of a weakly mixing rationally ergodic rank-one

transformation that is not rationally weakly mixing. We also mention an example of a weakly mixing, rationally ergodic, and Koopman mixing (or zero type, see §6.3 for the definition) transformation that is not subsequence rationally weakly mixing (Aaronson 2016).

The set of transformations that are subsequence rationally weakly mixing is residual (Aaronson 2013), while the set of rationally weakly mixing transformations is meagre (Aaronson 2013). Since the set of power weakly mixing rank-one transformations is residual, and the rank-one transformations are subsequence boundedly rationally ergodic, there exist rank one transformations that are power weakly mixing and subsequence boundedly rationally ergodic but not rationally weakly mixing.

Mixing, Zero-Type

We now consider several attempts to define (strong) mixing for nonsingular maps. Probably the first notion of mixing for infinite measure-preserving systems was proposed by Hopf in (1937). The idea was to show an asymptotic rate for the sequence $\mu(A \cap T^{-n}B)$ for a large class of finite measure sets A, B . More precisely, a transformation T is *mixing for a ring $\mathcal{R}$* (called now *Krickeberg mixing*), where $\mathcal{R}$ is a ring of sets of finite measure that is invariant under T and generates the entire σ -algebra, if there is a sequence $(\rho_n)_{n=1}^\infty$ such that for all $A, B \in \mathcal{R}$ we have

$$\lim_{n \rightarrow \infty} \rho_n \mu(A \cap T^{-n}B) = \mu(A)\mu(B).$$

Hopf proved such a property for an infinite measure-preserving transformation defined on $\mathbb{R}^+ \times [0, 1]$ that is now called an infinite random walk; with $\mathcal{R}$ being the ring of Riemann measurable subsets. If $\mathcal{R}$ is the ring of all subsets of finite measure, then there are no Krickeberg $\mathcal{R}$ -mixing transformations because of the existence of weakly wandering sets. We note that the above (purely measure theoretical) definition $\mathcal{R}$ -mixing is due to Friedman (1978) who extended Krickeberg's one (Krickeberg 1967) given for continuous transformations of topological spaces endowed with a measure. Recently, there have

been several works showing this version of mixing and computing mixing rates for several transformations. Melbourne and Terhesiu (Melbourne and Terhesiu 2012) have verified mixing for a large class of maps including AFN maps with indifferent fixed points; these methods were extended to invertible transformations by Melbourne (2015) and to additional maps by Gouëzel (2011). Recently, Dolgopyat and Nándori (2019) have shown Krickerberg mixing for a class of special flows; other recent work appeared in Bruin et al. (2019).

Another approach to mixing was proposed by Krengel and Sucheston (1969) for nonsingular maps. Given a sequence of measurable sets $\{A_n\}$, let $\sigma_k(\{A_n\})$ denote the σ -algebra generated by $A_k, A_{k+1}, \dots$. A sequence $\{A_n\}$ is said to be *remotely trivial* if $\bigcap_{k=0}^{\infty} \sigma_k(\{A_n\}) = \{\emptyset, X\} \bmod \mu$, and it is *semiremotely trivial* if every subsequence contains a further subsequence that is remotely trivial. A nonsingular transformation T of a σ -finite measure space is called *mixing* if for every set A of finite measure the sequence $\{T^{-n}A\}$ is semiremotely trivial, and *completely mixing* if $\{T^{-n}A\}$ is semiremotely trivial for all measurable sets A . Krengel and Sucheston show that T is completely mixing if and only if it is type II_1 and mixing for the equivalent finite invariant measure. Thus, there are no type III and II_{∞} completely mixing nonsingular transformations on probability spaces. We note that this definition of mixing in infinite measure spaces depends on the choice of measure inside the equivalence class (but it is independent if we replace the measure by an equivalent measure with the same collection of sets of finite measure).

Hajian and Kakutani showed (1964) that an ergodic infinite measure-preserving transformation T is either of *zero type*: $\lim_{n \rightarrow \infty} \mu(T^{-n}A \cap A) = 0$ for all sets A of finite measure, or of *positive type*: $\limsup_{n \rightarrow \infty} \mu(T^{-n}A \cap A) > 0$ for all subsets A of finite positive measure. It appears that T is mixing if and only if it is of zero type (Krengel and Sucheston 1969). We note that in infinite measure, mixing implies mixing of all orders, i.e., if a measure-preserving T is of zero type, then $\mu(T^{n_1}A_1 \cap \dots \cap T^{n_k}A_k) \rightarrow 0$ for each k and all

subsets $A_1, \dots, A_k$ of finite measure whenever $|n_i - n_j| \rightarrow \infty$ if $i \neq j$.

For $0 \leq \alpha \leq 1$, Kakutani suggested a related definition of α -type: An infinite measure-preserving transformation is of α -type if $\limsup_{n \rightarrow \infty} \mu(A \cap T^n A) = \alpha \mu(A)$ for every subset A of finite measure. In (Osikawa and Hamachi 1971), examples of ergodic transformations of any α -type and a transformation of not any type were constructed.

It was shown in Danilenko (2016a) and Loh et al. (2018) that for each pair $k \leq n$, there exists a mixing rank-one infinite measure-preserving transformation of ergodic index k and conservative index n . Rigid infinite measure-preserving rank-one transformations of arbitrary ergodic index were constructed in Danilenko (2016a). Of course, rigidity implies infinite conservative index.

We now isolate an important class of concrete rank-one transformations and examine mixing properties within this class. Let T be a rank-one transformation associated with a sequence $(r_n, w_n, s_n)_{n=1}^{\infty}$. If $w_n(0) = w_n(1) = \dots = w_n(r_n - 1)$ and $s_n(j) = z_n + j$ for $j = 0, \dots, r_n - 1$, then T is called a *high staircase* (called also tower staircase in Bowles et al. (2001)). It was shown in Bowles et al. (2001) that each high staircase is weakly doubly ergodic and hence weak mixing. However, there exist high staircases whose Cartesian square is not ergodic (Bowles et al. 2001). As for the mixing of the high staircases, the following theorem was proved in (Danilenko and Ryzhikov 2011). It is an infinite analogue of the Adams solution (Adams 1998) of the Smorodinsky conjecture.

Theorem 6.5 *If $\lim_{n \rightarrow \infty} \frac{r_n^2}{r_1 \dots r_{n-1}} = 0$ and $\sum_{n=1}^{\infty} \frac{z_n}{h_n} = \infty$, then the associated high staircase is infinite measure-preserving and mixing.*

Mixing high staircases which are power weakly mixing were constructed in (Danilenko and Ryzhikov 2011).

We note that mixing (zero type) does not imply either ergodicity or conservativity in the category of infinite measure-preserving transformations. Indeed, a translation on $\mathbb{R}$ endowed with the

Lebesgue measure is nonergodic, totally dissipative but of zero type. It may seem that mixing and ergodicity together are stronger than any kind of nonsingular weak mixing considered above. However, it is not the case: If T is a weakly mixing infinite measure-preserving transformation of zero type and S is an ergodic probability-preserving transformation, then $T \times S$ is ergodic and of zero type. On the other hand, the L^∞ -spectrum $e(T \times S)$ is nontrivial, i.e., $T \times S$ is not weakly mixing, whenever S is not weakly mixing. We also note that there exist rank-one infinite measure-preserving transformations T of zero type such that $T \times T$ is not conservative (hence not ergodic) (Adams et al. 1997). In contrast to that, if T is of positive type, then all of its finite Cartesian products are conservative (Aaronson and Nakada 2000). Another result that suggests that there is no good definition of mixing in the infinite measure-preserving case was proved in (James et al. 2008). It is shown there that while the mixing finite measure-preserving transformations are measurably sensitive, there exists no infinite measure-preserving system that is measurably sensitive. (Measurable sensitivity is a measurable version of the strong sensitive dependence on initial conditions – a concept from topological theory of chaos.)

The Krengel-Sucheston concept of mixing (or the Hajian-Kakutani zero type) considered above for infinite measure-preserving systems extends naturally to nonsingular transformations T of $(X, \mathcal{B}, \mu)$ without finite absolutely continuous invariant measure in the following way: We say that T is *Koopman mixing* (or of *zero type*) if the maximal spectral type of the Koopman operator U_T generated by T (see (3)) is a Rajchman measure, i.e., $\int_X f \cdot U_T^n f d\mu \rightarrow 0$ for each $f \in L^2(X, \mu)$. It is easy to see that this definition of mixing will not be affected if we replace μ with an equivalent measure. Examples of Koopman mixing rank-one transformations of type III were constructed in (Danilenko 2001a). Nonsingular Bernoulli transformations of type III are always Koopman mixing (Kosloff 2013).

More recently, Lenci (2010) introduced a new notion of mixing for infinite measure-preserving

maps that is motivated by statistical mechanics and uses global observables. The definition is with respect to a collection of sets, global observables, and local observables. We choose a family $\mathcal{V}$ of measurable sets of finite measure so that it contains sets $V_1 \subset V_2 \subset \dots$ such that $\cup_i V_i = X$. We also have a subspace $\mathcal{G}$ of L^∞ functions (called global observables), and a subspace $\mathcal{L}$ of L^1 functions (called local observables). There is also a condition on the growth rate of the measure of $\mathcal{V}$ -elements under iteration by T . Then Lenci defines an *infinite volume average* for elements F of $\mathcal{G}$ by

$$\bar{\mu}(F) = \lim_{V \rightarrow X} \int_V F d\mu.$$

By this limit, we mean that for every neighborhood of $\bar{\mu}(F)$, there is a number $M > 0$ so that when $\mu(V) > M$ for a set V in $\mathcal{V}$, then $\int_V F d\mu$ is in the neighborhood. He shows that under the above conditions, $\bar{\mu}(F \circ T^n) = \bar{\mu}(F)$. Then he defines several notions of what he calls infinite volume mixing (Lenci 2010); we mention three here. The transformation T is said to be *global-local mixing-1* if for all F in $\mathcal{G}$ and all g in $\mathcal{L}$ with $\int g d\mu = 0$, we have

$$\lim_{n \rightarrow \infty} \int (F \circ T^n) g d\mu = 0.$$

The transformation T is said to be *global-local mixing-2* if for all F in $\mathcal{G}$ and all g in $\mathcal{L}$ we have

$$\lim_{n \rightarrow \infty} \int (F \circ T^n) g d\mu = \bar{\mu}(F) \int g d\mu.$$

The transformation is said to be *global-global mixing* if for all F, G in $\mathcal{G}$ we have

$$\lim_{n \rightarrow \infty} \bar{\mu}(F \circ T^n G) d\mu = \bar{\mu}(F) \bar{\mu}(G)$$

Lenci proves in (Lenci 2013) that if T is an infinite measure-preserving K-automorphism, then T is global-local mixing-1 for any choice of $\mathcal{V}$ satisfying the measure growth condition, for $\mathcal{L} = L^1$, and for C that is the closure in L^1 of $T^n \mathcal{F}$, where $\mathcal{F}$ is as in the definition of the

K -automorphism (see subsection “ K -automorphism”). Infinite mixing has been shown for other examples (Lenci 2010), in particular for uniformly expanding maps of the interval (Lenci 2017), and for one-dimensional maps with an indifferent fixed point (Bonanno et al. 2018).

K -property

A nonsingular transformation T of $(X, \mathcal{B}, \mu)$ is called K -automorphism (Silva and Thieullen 1995) if there exists a sub- σ -algebra $\mathcal{F} \subset \mathcal{B}$ such that $T^{-1}\mathcal{F} \subset \mathcal{F}$, $\bigcap_{k \geq 0} T^{-k}\mathcal{F} = \{\emptyset, X\}$, $\bigvee_{k=0}^{+\infty} T^k\mathcal{F} = \mathcal{B}$ and the Radon-Nikodym derivative ω_1^μ is $\mathcal{F}$ -measurable (see also Parry 1965) for the case when T is of type II_∞ ; the authors in (Silva and Thieullen 1995) required T to be conservative). If R is a nonsingular endomorphism on $(X, \mathcal{B}, \mu)$, then the natural extension of R is a K -automorphism if and only if R is exact, i.e., $\bigwedge_{n=1}^\infty R^{-n}\mathcal{B} = \{\emptyset, X\} \bmod 0$. It follows from the Kolmogorov 0–1 theorem that a nonsingular Bernoulli shift from the generalized Krengel class (see § 3.5) is a K -automorphism. Parry (Parry 1965) showed that a type II_∞ K -automorphism is either dissipative or ergodic. Krengel (1970) proved the same for the Krengel class of Bernoulli nonsingular shifts, and finally Silva and Thieullen extended this result to the nonsingular K -automorphisms (Silva and Thieullen 1995). It is also shown in (Silva and Thieullen 1995) that if T is a nonsingular K -automorphism, for any ergodic nonsingular transformation S , if $S \times T$ is conservative, then it is ergodic. This property is called *sharply weak mixing* in (Danilenko and Lemańczyk 2022). It follows that a conservative nonsingular K -automorphism is weakly mixing. However, it does not necessarily have infinite ergodic index (Kakutani and Parry 1963). Krengel and Sucheston (1969) showed that an infinite measure-preserving conservative K -automorphism is mixing. In (Elyze et al. 2018), it is shown that if T is a rank-one transformation with (r_n) bounded and $s_{n-1}(r_{n-1} - 1) \geq h_n/2$ for all $n \geq 1$ (so the space is of infinite measure, and T is the infinite Chacón map), there exists a rank-one transformation S such that $T \times S$ is conservative but not ergodic, so it is not sharply weak mixing.

Multiple and Polynomial Recurrence

Let p be a positive integer. A nonsingular transformation T is called p -recurrent if for every subset B of positive measure there exists a positive integer k such that

$$\mu(B \cap T^{-k}B \cap \dots \cap T^{-kp}B) > 0.$$

If T is p -recurrent for any $p > 0$, then it is called *multiply recurrent*. It is easy to see that T is 1-recurrent if and only if it is conservative. Clearly, if T is rigid, then it is multiply recurrent. Furstenberg showed (Furstenberg 1981) that every finite measure-preserving transformation is multiply recurrent. In contrast to that, Eigen, Hajian, and Halverson (Eigen et al. 1998a) constructed for any $p \in \mathbb{N} \cup \{\infty\}$, a nonsingular product odometer of type II_∞ which is p -recurrent but not $(p+1)$ -recurrent. Aaronson and Nakada showed in (2000) that an infinite measure-preserving Markov shift T is p -recurrent if and only if the product $T \times \dots \times T$ (p times) is conservative. It follows from this and (Aaronson et al. 1979) that in the class of ergodic Markov shifts, infinite ergodic index implies multiple recurrence. However, in general this is not true. It was shown in (Adams et al. 2001; Gruher et al. 2003; Danilenko and Silva 2004) that for each $p \in \mathbb{N} \cup \{\infty\}$ there exist

- (i) Power weakly mixing rank-one transformations
- (ii) Nonpower weakly mixing rank-one transformations with infinite ergodic index

which are p -recurrent but not $(p+1)$ -recurrent (the latter holds when $p \neq \infty$, of course).

A subset A is called p -wandering if $\mu(A \cap T^k A \cap \dots \cap T^{pk} A) = 0$ for each k . Aaronson and Nakada established in (Aaronson and Nakada 2000) a p -analogue of Hopf decomposition (see Theorem 2.2).

Proposition 6.6 *If $(X, \mathcal{B}, \mu, T)$ is conservative aperiodic nonsingular dynamical system and $p \in \mathbb{N}$, then $X = C_p \sqcup D_p$, where C_p and D_p are T -invariant disjoint subsets, D_p is a countable*

union of p -wandering sets, $T \upharpoonright C_p$ is p -recurrent, and $\sum_{k=1}^{\infty} \mu(B \cap T^{-k}B \cap \dots \cap T^{-dk}B) = \infty$ for every $B \subset C_p$.

Let T be an infinite measure-preserving transformation and let $\mathcal{F}$ be a σ -finite factor (i.e., invariant subalgebra) of T . Inoue (Inoue 2004) showed that for each $p > 0$, if $T \upharpoonright \mathcal{F}$ is p -recurrent, then so is T provided that the extension $T \rightarrow T \upharpoonright \mathcal{F}$ is isometric. It is unknown yet whether the latter assumption can be dropped. However, partial progress was achieved in (Meyerovitch 2007): If $T \upharpoonright \mathcal{F}$ is multiply recurrent, then so is T .

Let $\mathcal{P} := \{q \in \mathbb{Q}[t] \mid q(\mathbb{Z}) \subset \mathbb{Z} \text{ and } q(0) = 0\}$. An ergodic conservative nonsingular transformation T is called p -polynomially recurrent if for every $q_1, \dots, q_p \in \mathcal{P}$ and every subset B of positive measure there exists $k \in \mathbb{N}$ with

$$\mu(B \cap T^{q_1(k)}B \cap \dots \cap T^{q_p(k)}B) > 0.$$

If T is p -polynomially recurrent for every $p \in \mathbb{N}$, then it is called *polynomially recurrent*. Furstenberg's theorem on multiple recurrence was significantly strengthened in (Bergelson and Leibman 1996), where it was shown that every finite measure-preserving transformation is polynomially recurrent. However, Danilenko and Silva (Danilenko and Silva 2004) constructed

- (i) Type II_{∞} transformations T which are p -polynomially recurrent but not $(p+1)$ -polynomially recurrent (for each fixed $p \in \mathbb{N}$)
- (ii) Polynomially recurrent transformations T of type II_{∞} ,
- (iii) Rigid (and hence multiply recurrent) type II_{∞} transformations T which are not polynomially recurrent

Moreover, such T can be chosen inside the class of rank-one transformations with infinite ergodic index.

Dynamical Properties of IDPFT Systems

Let T_n be an ergodic transformation of a standard probability space (X_n, μ_n, T_n) , and let ν_n be a μ_n -

equivalent invariant probability measure for each $n \in \mathbb{N}$. Assume that the transformation $T = \bigotimes_{n=1}^{\infty} T_n$ is nonsingular with respect to the infinite product measure $\mu := \bigotimes_{n=1}^{\infty} \mu_n$. In other words, (X, μ, T) is an IDPFT dynamical system.

Theorem 7.1 *The transformation T is either conservative or totally dissipative. Moreover, if S is an ergodic conservative nonsingular transformation, then the direct product $T \times S$ is either conservative or totally dissipative (Danilenko and Lemańczyk 2019).*

Theorem 7.2 *Let (X_n, ν_n, T_n) be mildly mixing (see §15.1 for the definition) for each $n > 0$. If T is μ -conservative, then T is sharply weak mixing (Danilenko and Lemańczyk 2019).*

Examples of rigid ergodic but not weakly mixing IDPFT transformations of Krieger type III_{λ} , for each $\lambda \in (0, 1)$ were constructed in (Danilenko and Lemańczyk 2019). Some families of 0-type IDPFT transformations of type III_1 appeared in (Danilenko and Lemańczyk 2019) and of all possible Krieger types in (Danilenko and Kosloff 2022).

Theorem 7.3 *Let $K \in \{III_{\lambda} \mid 0 \leq \lambda \leq 1\} \sqcup \{II_{\infty}\}$. Then there is a 0-type weakly mixing IDPFT transformation of type K (Danilenko and Kosloff 2022).*

Dynamical Properties of Nonsingular Bernoulli and Markov Shifts

We will use below the notation introduced in §3.5. Thus, T stands for the nonsingular Bernoulli shift on the space $(X, \mu) = (A^{\mathbb{Z}}, \bigotimes_{n \in \mathbb{Z}} \mu_n)$ associated with a sequence of probability measures $(\mu_n)_{n \in \mathbb{Z}}$ on A .

Krengel Class

Nonsingular Bernoulli shifts appeared originally in Krengel's work (Krengel 1970). He introduced there a class of nonsingular shifts for which $A = \{0, 1\}$ and μ_n is the equidistribution on $\{0, 1\}$ for all $n \leq 1$. We will call it *the Krengel class*. Krengel showed that this class contains totally dissipative transformations. He also used an inductive procedure to construct the sequence $(\mu_n)_{n=1}^{\infty}$ in such a way that the corresponding Bernoulli shift is

ergodic conservative and not of type II_1 . Krengel conjectured that the shift is of type III indeed. In (Hamachi 1981b), Hamachi showed that Krengel's class contains ergodic conservative nonsingular Bernoulli shifts of type III . This was further refined by Kosloff who constructed type III_1 ergodic conservative shifts belonging to Krengel's class (Kosloff 2011). In (Kosloff 2013), Kosloff constructed a nonsingular Bernoulli shift of type III_1 (and belonging to the Krengel class) which is power weakly mixing. Weiss asked about possible Krieger's types for the nonsingular Bernoulli shifts. Answering his question, Kosloff proved in a subsequent paper (Kosloff 2014) that each conservative Bernoulli shift from the Krengel class is ergodic and either of type II_1 or of type III_1 . In particular, the non-type- II_1 conservative Bernoulli shifts constructed in (Krengel 1970; Hamachi 1981b; Kosloff 2011)) are all of type III_1 indeed.

Generalized Krengel Class

Kosloff's result from (Kosloff 2014) was further extended in (Danilenko and Lemańczyk 2019). We say that a nonsingular Bernoulli shift belongs to the *generalized Krengel class* if $A = \{0, 1\}$ and $\mu_n = \mu_1$ for each $n \leq 0$. We note that these transformations are the natural extension of the one-sided nonsingular Bernoulli shifts defined on $(A^{\mathbb{N}}, \otimes_{n>0} \mu_n)$. Every shift from the generalized Krengel class is a K -automorphism.

Theorem 8.1 (On types of nonsingular Bernoulli shifts from the generalized Krengel class (Kosloff 2014; Danilenko and Lemańczyk 2019)). *Let $A = \{0, 1\}$ and let T be a nonsingular Bernoulli shift on $(A^{\mathbb{Z}}, \otimes_{n \in \mathbb{Z}} \mu_n)$ from the generalized Krengel class.*

- (i) *If $\sum_{n>0} (\mu_n(0) - \mu_1(0))^2 < \infty$, then μ is equivalent to $\otimes_{n \in \mathbb{Z}} \mu_1$ and hence T is of type II_1 .*
- (ii) *If $\sum_{n>0} (\mu_n(0) - \mu_1(0))^2 = \infty$ and T is conservative, then T is ergodic of type III_1 . Moreover, the Maharam extension of T is a weakly mixing K -automorphism.*

Thus, Krieger's type of each nonsingular Bernoulli shift from the generalized Krengel

class is never of type III_λ , $0 \leq \lambda < 1$. In (Vaes and Wahl 2018), Vaes and Wahl, answering a question from (Danilenko and Lemańczyk 2019), found a convenient condition for a nonsingular Bernoulli shift from the generalized Krengel class to be conservative. Utilizing that condition, they constructed, for each $\lambda \in (0, 1)$, an explicit example of power weakly mixing nonsingular Bernoulli shift of type III_1 with $\mu_1(0) = \lambda$. We note that the previously known Bernoulli shifts of type III_1 were constructed via involved inductive procedures. Vaes and Wahl also provided in (Vaes and Wahl 2018) a family of type III_1 -examples of Bernoulli shifts that contains examples with finite ergodic index (less than 73). Analyzing that family more thoroughly, Kosloff and Soo showed that it contains type III_1 nonsingular Bernoulli shifts of arbitrary ergodic index.

Theorem 8.2 *Let $A = \{0, 1\}$ and $c > 0$. Let $\mu_n^c(0) := \frac{1}{2} + \frac{c}{\sqrt{n}} 1_{\{n \in \mathbb{N} \mid \sqrt{n} > 2c\}}$ for each $n \in \mathbb{Z}$. There exists $D > \frac{1}{6}$ such that the Bernoulli shift T on $\otimes_{n \in \mathbb{Z}} (A, \mu_n)$ is ergodic of type III_1 for all $c < D$ and totally dissipative for all $c > D$. In addition, if $k \in \mathbb{N}$ and $\frac{D}{\sqrt{k+1}} \leq c < \frac{D}{\sqrt{k}}$ then T is of ergodic index k (Kosloff and Soo 2022).*

General Nonsingular Bernoulli Shifts

The studying of general nonsingular Bernoulli shifts was initiated by Kosloff in (2013).

Theorem 8.3 (Mixing of nonsingular Bernoulli shifts (Kosloff 2013)). *If $\#A = 2$ and $(\mu_n)_{n \in \mathbb{Z}}$ is a sequence of probabilities on A such that (4) holds, then T is either of type II_1 and mixing (with respect to the equivalent invariant probability measure) or of zero type.*

In (Kosloff 2019), Kosloff noticed that under some natural conditions, conservativity of Bernoulli shifts implies ergodicity. His proof was based essentially on the Hurewicz ergodic theorem and properties of the tail equivalence relation on $A^{\mathbb{Z}}$. Danilenko (Danilenko 2019a) refined his results by exploiting the interplay between T and the measurable equivalence relation on $A^{\mathbb{Z}}$ generated by the finite permutations of coordinates.

Theorem 8.4 (Weak mixing of conservative nonsingular Bernoulli shifts). *Let A be finite.*

- (i) *If $\inf_{n \in \mathbb{Z}} \min_{a \in A} \mu_n(a) > 0$ and T is conservative, then T is weakly mixing (see Kosloff 2019; Danilenko 2019a).*
- (ii) *If $\#A = 2$ and $\inf_{n \in \mathbb{Z}} \min_{a \in A} \log \left| \frac{\mu_n(a)}{\mu_{n+1}(a)} \right| > 0$ and T is conservative, then T is weakly mixing (Danilenko 2019a).*
- (iii) *Under condition (i) or (ii), if $T \times \cdots \times T$ (p times) is conservative for some $p \geq 1$, then $T \times \cdots \times T$ (p times) is weakly mixing (Danilenko 2019a).*

The techniques from (Kosloff 2019; Danilenko 2019a) were further elaborated in (Björklund et al. 2021) to prove the following refinement of Theorem 8.4.

Theorem 8.5 *Let $A = \{0, 1\}$ and let T be a nonsingular Bernoulli shift on $(A^{\mathbb{Z}}, \otimes_{n \in \mathbb{Z}} \mu_n)$ and let $\mu = \otimes_{n \in \mathbb{Z}} \mu_n$ be nonatomic. If T is not totally dissipative, then T is weakly mixing and its type is given as follows.*

- (i) *If there is $\lambda \in (0, 1)$ such that $\sum_{n \in \mathbb{Z}} (\mu_n(0) - \lambda)^2 < +\infty$, then T is of type II_1 .*
- (ii) *If there is $\lim_{n \rightarrow \infty} \mu_n(0) = \lambda \in (0, 1)$ and $\sum_{n \in \mathbb{Z}} (\mu_n(0) - \lambda)^2 = +\infty$, then T is of type III_1 .*
- (iii) *If either $\lim_{n \rightarrow \infty} \mu_n(0) = 0$ or $\lim_{n \rightarrow \infty} \mu_n(0) = 1$, then T is of type III .*
- (iv) *If the sequence $(\mu_n(0))_n$ does not converge as $n \rightarrow \infty$, then T is of type III_1 .*

It follows, in particular, that the nonsingular Bernoulli shifts on $\{0, 1\}^{\mathbb{Z}}$ are never of type II_{∞} . It is still an open problem which subtypes of III are realized in the case (iii). An example of T of type III_1 with $\lim_{n \rightarrow \infty} \mu_n(0) = 0$ was constructed in (Berendschot and Vaes 2022a). Is there T of type III_0 satisfying (iii)?

Parts (i) and (ii) of Theorem 8.5 were generalized to any finite set A in Avraham-Re'em (2022, Remark 8.7).

The situation is different if we consider $A = [0, 1]$ and the probability measures $(\mu_n)_{n \in \mathbb{Z}}$ on A with infinite support. In (Kosloff and Soo 2021), for

each $\lambda \in (0, 1)$, an example of type III_{λ} Bernoulli shift was constructed with $\mu_n \sim \text{Leb}$ for each n . Examples of nonsingular Bernoulli shifts of each possible Krieger's type were given in a later paper (Berendschot and Vaes 2022a). An alternative proof of this result appeared in a recent work (Danilenko and Kosloff 2022).

Theorem 8.6 *Let $K \in \{III_{\lambda} \mid 0 \leq \lambda \leq 1\} \sqcup \{II_{\infty}\}$. Then there is a sequence $(\mu_n)_{n \in \mathbb{Z}}$ of probability measures on $[0, 1]$ such that $\mu_n \sim \text{Leb}$ for each $n \in \mathbb{N}$, the Bernoulli shift T on $([0, 1]^{\mathbb{Z}}, \otimes_{n \in \mathbb{Z}} \mu_n)$ is weakly mixing, IDPFT, and of Krieger type K (Danilenko and Kosloff 2022).*

We also note that if there is $C > 1$ such that $C^{-1} \leq \frac{d\mu_n}{d\mu_0} \leq C$ for all $n \in \mathbb{Z}$, then T can be neither of type III_0 nor of type II_{∞} (Berendschot and Vaes 2022a).

The following result is an analog of Theorem 8.5 for Bernoulli shifts with general state spaces.

Theorem 8.7 *Let $(\mu_n)_{n \in \mathbb{Z}}$ be a sequence of mutually equivalent probabilities on a standard Borel space A and let T be the Bernoulli shift on $(X, \mu) = \otimes_{n \in \mathbb{Z}} (A, \mu_n)$. If T is nonsingular and conservative, then T is weakly mixing and the following holds (Berendschot and Vaes 2022b).*

- (i) *T of type II_1 if and only if there exists a probability measure $\nu \sim \mu_0$ such that $\nu^{\otimes \mathbb{Z}} \sim \mu$.*
- (ii) *T of type II_{∞} if and only if there exists a σ -finite measure $\nu \sim \mu_0$ and Borel subsets $B_n \subset A$ such that $\nu(B_n) < \infty$ for each $n \in \mathbb{Z}$ and*

$$\sum_{n \in \mathbb{Z}} \mu_n(A \setminus B_n) < \infty, \sum_{n \in \mathbb{Z}} \nu_n(A \setminus B_n) = \infty,$$

$$\sum_{n \in \mathbb{Z}} H^2 \left(\mu_n, \nu(B_n)^{-1} \nu \upharpoonright B_n \right)$$

- (iii) *T is of type III in all other cases.*

The next very natural question to ask is: Which ergodic flows arise as the associated flow of nonsingular Bernoulli shifts of type III_0 ?

Definition 8.8 Let $V = (V_t)_{t \in \mathbb{R}}$ be a nonsingular flow on a standard probability space (X, ν) . Given $n > 1$, consider two mutually commuting nonsingular flows $U = (U_t)_{t \in \mathbb{R}}$ and $D = (D_t)_{t \in \mathbb{R}}$ on the product space $(X, \nu)^{\otimes n}$:

$$\begin{aligned} U_t(x_1, \dots, x_n) &:= (V_t x_1, x_2, \dots, x_n), \\ D_t(x_1, \dots, x_n) &:= (V_t x_1, V_t x_2, \dots, V_t x_n), \end{aligned}$$

for each $t \in \mathbb{R}$ and $(x_1, \dots, x_n) \in X^n$. The restriction of U to the σ -algebra of D -invariant subsets in X^n is a well-defined nonsingular flow. It is called the *joint flow of n copies of V* . The flow V is called *infinitely divisible* if for each $n > 1$, there is a flow W such that V is isomorphic to the joint flow of n copies of W .

If V is an AT-flow associated with an ITPFI₂-transformation (see section “ITPFI Transformations and AT-Flows”), then V is infinitely divisible.

Theorem 8.9

- (i) Let $(\mu_n)_{n \in \mathbb{Z}}$ be a sequence of mutually equivalent probabilities on a standard Borel space A , and let T be the Bernoulli shift on $\otimes_{n \in \mathbb{Z}} (A, \mu_n)$. If T is nonsingular and conservative, then the associated flow of T is infinitely divisible. For each ergodic probability-preserving transformation S , the product $T \times S$ is ergodic and the associated flows of T and $T \times S$ are isomorphic (Berendschot and Vaes 2022b).
- (ii) Conversely, let V be an AT-flow associated with an ITPFI₂-transformation. Then there is a weakly mixing nonsingular Bernoulli shift whose associated flow is isomorphic to V .

We note that formally Berendschot and Vaes proved in (Berendschot and Vaes 2022b) a stronger claim than (ii). They introduced a concept of *Poisson flow* as the tail boundary of certain non-homogeneous random walks on $\mathbb{R}$ (see §15.7 below). Every Poisson flow is infinitely divisible. Every AT-flow associated with an ITPFI₂-transformation is Poisson. It is shown in (Berendschot and Vaes 2022b) that Theorem 8.9 (ii) holds for each Poisson flow V . However, the following problems remain open:

- Whether each Poisson flow is an AT-flow associated with an ITPFI₂-transformation.
- Whether each infinitely divisible flow is Poisson, whether each infinitely divisible AT-flow is Poisson.

Bernoulli Factors of Nonsingular Bernoulli Shifts

Definition 8.10 Let T be a nonsingular Bernoulli shift on $(X, \mu) := \otimes_{n \in \mathbb{Z}} (A, \mu_n)$. Suppose that there is another sequence $(\nu_n)_{n \in \mathbb{Z}}$ of probabilities on A such that T is nonsingular on the product space $(X, \nu) := \otimes_{n \in \mathbb{Z}} (A, \nu_n)$. If there is a map $\pi: A^{\mathbb{Z}} \rightarrow A^{\mathbb{Z}}$ such that $\pi T = T\pi$ and the measure $\mu \circ \pi^{-1}$ is equivalent to ν , then (X, ν, T) is called a *factor* of (X, μ, T) .

We assume that the factors are nontrivial. What is the relation between Krieger’s types of (X, ν, T) and (X, μ, T) ?

Theorem 8.11 (Bernoulli factors of type II₁ (Kosloff and Soo 2022)) Let A be finite and

$$1 > \inf_{n \in \mathbb{Z}} \min_{a \in A} \mu_n(a) > 0.$$

Then there is a probability measure ρ on A such that $(X, \rho^{\otimes \mathbb{Z}}, T)$ is a factor of (X, μ, T) .

Theorem 8.12 Let $\lambda, \lambda' \in (0, 1]$. There exists a type III _{λ} Bernoulli shift which has a type III _{λ'} Bernoulli shift as a factor in each of the two cases (Kosloff and Soo 2022):

- (i) $\lambda' < \lambda = 1$,
- (ii) $\lambda < \lambda'$.

Nonsingular Markov Shifts

An analog of Theorem 8.4(i) holds also for nonsingular Markov shifts (see (Kosloff 2019; Danilenko 2019a)).

Theorem 8.13 (Weak mixing of conservative nonsingular Markov shifts (Kosloff 2019; Danilenko 2019a)). Let A be finite, and let $M = (M(a, b))_{a, b \in A}$ be a 0–1-valued $A \times A$ -matrix. Suppose that M is primitive, i.e., there is $n > 0$

such that all the entries of M^n are strictly positive. Let (X_M, T, μ) be a nonsingular Markov shift, and let μ be generated by a sequence $(\pi_n, P_n)_{n \in \mathbb{Z}}$ as in §3.6. Suppose that μ is nonatomic and that π_n is fully supported on A for each n . If $\inf \{P_n(a, b) \mid n \in \mathbb{Z}, M(a, b) = 1\} > 0$ and T is conservative, then T is weakly mixing.

We isolate a class of nonsingular Markov shifts for which $P_n = P_1$ and $\pi_n = \pi_1$ for all $n \leq 0$ and call it the Markov-Krengel class. Each shift from this class is the natural extension of the corresponding one-sided nonsingular Markov shift (Danilenko and Lemańczyk 2019). There is an analog of Theorem 8.1 for the Markov-Krengel shifts.

Theorem 8.14 (Danilenko and Lemańczyk 2019). Let $M^* = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, P_n be a bistochastic matrix for each $n \in \mathbb{Z}$, $P_k = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$ and $\pi_k = (0.5, 0.5)$ for each $k \leq 0$. Let the corresponding Markov-Krengel shift (X_{M^*}, T, μ) be nonsingular and conservative. Then either T is of type II_1 (if $\sum_{n>0} |P_n(0, 0) - 0.5| < \infty$) or III_1 (otherwise). In the latter case, the Maharam extension of T is a weakly mixing K -automorphism. Moreover, if μ is equivalent to a Bernoulli (i.e., infinite product) measure, then T is of type II_1 .

Concrete examples of Markov-Krengel shifts (X_M, μ, T) of type III_1 such that μ is not equivalent to a Bernoulli measure were constructed in (Danilenko and Lemańczyk 2019; Kosloff 2021).

Recently, Avraham-Re'em extended and refined the aforementioned results on Markov shifts (Avraham-Re'em 2022). To state his results, we introduce some notation. Given $n \in \mathbb{Z}$ and $a, b \in A$, we let

$$\hat{P}_n(a, b) = \begin{cases} \frac{\pi_{n-1}(b)}{\pi_n(a)} P_{n-1}(b, a) & \text{if } \pi_n(a) \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

For a stochastic $A \times A$ -matrix Q , let λ be the distribution on A such that $\lambda Q = \lambda$. We let $\hat{Q}(a, b) := \frac{\lambda(b)}{\lambda(a)} Q(b, a)$.

Theorem 8.15 Let A be finite and let $M = (M(a, b))_{a, b \in A}$ be a primitive 0–1-valued $A \times A$ -matrix. Let a measure μ on X_M be generated by a sequence $(\pi_n, P_n)_{n \in \mathbb{Z}}$ as in §3.6 and $\inf \{P_n(a, b) \mid n \in \mathbb{Z}, M(a, b) = 1\} > 0$. Let the Markov shift (X_M, T, μ) be nonsingular and conservative (Avraham-Re'em 2022).

- (i) If $\lim_{n \rightarrow \infty} P_n$ does not exist, then T is of type III_1 .
- (ii) If there exists the limit $P_+ := \lim_{n \rightarrow +\infty} P_n$ and $P_- := \lim_{n \rightarrow -\infty} P_n$, then $P_+ = P_-$.
- (iii) If $A = \{0, 1\}$ and there exists the limit $Q := \lim_{n \rightarrow \infty} P_n$, then T is either of type II_1 or III_1 . More precisely, T is of type II_1 if and only if

$$\sum_{n \geq 1} \sum_{a, b, a', b' \in A} \left(\sqrt{\hat{P}_{-n}(a, b) P_n(a', b')} - \sqrt{\hat{Q}(a, b) Q(a', b')} \right)^2 < \infty.$$

The corresponding equivalent invariant probability measure (if exists) is the Markov measure defined by Q and the distribution λ on A satisfying $\lambda Q = \lambda$.

It was also shown in (Avraham-Re'em 2022) that an analogue of Theorem 8.15(iii) holds also for the golden mean Markov shift for which $A =$

$$\{0, 1, 2\} \text{ and } M = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Dynamical Properties of Nonsingular Poisson Suspensions and Nonsingular Gaussian Transformations

Nonsingular Poisson Suspensions

Let $(X, \mathcal{B}, \mu)$ be a σ -finite infinite standard measure space and let $T \in \text{Aut}_2(X, \mu)$. Then the Poisson suspension (X^*, μ^*, T_*) is well defined by Theorem 3.1. The first problem to consider is to find out when T_* admits an absolutely continuous invariant probability measure. A satisfactory solution of this problem is obtained in (Danilenko et al. 2022a).

Theorem 9.1 *The following are equivalent:*

- *There exists a T_* -invariant probability measure $\rho \prec \mu_*$,*

$$\sup_{n \in \mathbb{Z}} \left\| \sqrt{\frac{d\mu \circ T^n}{d\mu}} - 1 \right\|_2 < \infty,$$

- *There is a measure $\kappa \prec \mu$ such that $\sqrt{\frac{d\kappa}{d\mu}} - 1 \in L^2(\mu)$ and $\kappa \circ T = \kappa$.*

It follows that there exists “properly” non-singular Poisson suspensions, i.e., suspensions that do not admit an absolutely continuous invariant measure. The next problem is to find out when T_* is dissipative and when it is conservative.

Theorem 9.2 (i) *If $\sum_{n \in \mathbb{Z}} e^{-\frac{1}{2} \left\| \sqrt{\frac{d\mu \circ T^n}{d\mu}} - 1 \right\|_2^2} < \infty$, then T_* is totally dissipative (Danilenko et al. 2022a).*

(ii) *Let $T \in \text{Aut}_1(X, \mu)$, $\chi(T) = 0$ and $\left(\frac{d\mu}{d\mu \circ T^n} \right)^2 - 1 \in L^1(X, \mu)$ for each $n \in \mathbb{N}$. If there is a sequence $(b_n)_{n=1}^\infty$ of positive reals such that $\sum_{n=1}^\infty b_n = \infty$ but*

$$\sum_{n=1}^\infty b_n^2 e^{\int_X \left(\left(\frac{d\mu}{d\mu \circ T^n} \right)^2 - 1 \right) d\mu} < \infty,$$

then T_* is conservative (Danilenko et al. 2022b).

We also note that if $T \in \text{Aut}_1(X, \mu)$ and $\chi(T) \neq 0$, then T is not conservative (Danilenko et al. 2022a). Moreover, T_* is totally dissipative (Danilenko et al. 2022b).

Theorem 9.3 *If T is of 0-type and there is no T -invariant measure $\kappa \prec \mu$ such that $\sqrt{\frac{d\kappa}{d\mu}} - 1 \in L^2(\mu)$, then T_* is of 0-type (Danilenko et al. 2022a).*

The following theorem is proved via Baire category tools.

Theorem 9.4 *The set*

$$\{T \in \text{Aut}_2(X, \mu) \mid T \text{ and } T_* \text{ are both ergodic and of type III}_1\}$$

is a dense G_δ in $(\text{Aut}_2(X, \mu), d_2)$. The set

$$\{T \in \ker \chi \mid T \text{ and } T_* \text{ are both ergodic and of type III}_1\}$$

is a dense G_δ in $(\ker \chi, d_1)$. (Danilenko et al. 2022b).

It is easy to see that if $T \in \text{Aut}_2(X, \mu)$, then $T \in \text{Aut}_2(X, t \cdot \mu)$ for each $t > 0$. Hence T_* is $(t \cdot \mu)^*$ -nonsingular for each $t > 0$. However the dynamical properties of the systems $(X^*, T_*, (t \cdot \mu)^*)$ depend heavily on the choice of t . This is illustrated by a concrete example constructed in (Danilenko et al. 2022b).

Example 9.5 (Phase transition). There is a totally dissipative transformation $T \in \text{Aut}_2(X, \mu)$ and $t_0 > 0$ such that the dynamical system $(X^*, T_*, (t \cdot \mu)^*)$ is ergodic of type III₁ for each $t \in (0, t_0)$ and totally dissipative for each $t > t_0$.

The point t_0 can be interpreted as a “bifurcation point.”

We note that if $T \in \text{Aut}_2(X, \mu)$ and T is totally dissipative, then T_* is isomorphic to a nonsingular Bernoulli shift. Therefore, the following theorem was proved in (Danilenko and Kosloff 2022) simultaneously with Theorem 8.6.

Theorem 9.6 *Let $K \in \{\text{III}_\lambda \mid 0 \leq \lambda \leq 1\} \sqcup \{\text{II}_\infty\}$. Then there is a totally dissipative transformation $T \in \text{Aut}_2(X, \mu)$ such that (X^*, T_*, μ^*) is weakly mixing of type K .*

This theorem is further refined in type III₀ as follows (cf. Theorem 8.9).

Theorem 9.7 *Let V be an AT-flow associated with an ITPFI₂-transformation. Then there is a totally dissipative transformation $T \in \text{Aut}_2(X, \mu)$ such that T_* is weakly mixing and the associated flow of T_* is isomorphic to V (Berendschot and Vaes 2022b).*

Nonsingular Gaussian Transformations

Let $\mathcal{H}$ be a separable infinite dimensional real Hilbert space. Let $(h, O) \in \text{Aff } \mathcal{H}$. For $n \in \mathbb{Z}$, define a vector $h^{(n)} \in \mathcal{H}$ by setting $(h, O)^n = (h^{(n)}, O^n)$. We first consider when $G_{h, O}$ admits an equivalent invariant probability measure.

Theorem 9.8 *Let (X, μ) denote the space of $G_{h, O}$. The following are equivalent (Arano et al. 2021; Danilenko and Lemańczyk 2022):*

- *There exists a $G_{h, O}$ -invariant probability measure $\rho \sim \mu$,*
- *h is an O -coboundary, i.e., there is $a \in \mathcal{H}$ such that $h = a - Oa$,*
- *The affine operator $(h, O) \in \text{Aff } \mathcal{H}$ on $\mathcal{H}$ has a fixed point.*
- *The sequence $(h^{(n)})_{n \in \mathbb{Z}}$ is bounded in $\mathcal{H}$.*

The following dichotomy for nonsingular Gaussian transformations was established in (Arano et al. 2021) (see also (Danilenko and Lemańczyk 2022)).

Theorem 9.9 *If O has no nontrivial fixed vectors in $\mathcal{H}$ then $G_{h, O}$ is either conservative or totally dissipative.*

In (Arano et al. 2021), for each $(h, O) \in \text{Aff } \mathcal{H}$, the authors consider a one-parametric family of nonsingular Gaussian transformations $G_{\theta h, O}$, $\theta \in (0, +\infty)$.

Definition 9.10 *The Poincaré exponent of (h, O) is*

$$\delta_{h, O} := \inf \left\{ \alpha > 0 \mid \sum_{n=1}^{\infty} e^{-\alpha \|h^{(n)}\|^2} < \infty \right\} \in [0, +\infty].$$

The following theorem demonstrates a phase transition from conservativity to total dissipativity.

Theorem 9.11 *Let $(h, O) \in \text{Aff } \mathcal{H}$. There exists $\theta_{\text{diss}} \in [0, +\infty]$ such that $G_{\theta h, O}$ is conservative for all $\theta < \theta_{\text{diss}}$ and totally dissipative for all $\theta > \theta_{\text{diss}}$. Moreover,*

$$\sqrt{2\delta_{h, O}} \leq \theta_{\text{diss}} \leq 2\sqrt{2\delta_{h, O}}.$$

This result can be strengthened under additional assumptions on O (Arano et al. 2021).

Theorem 9.12 *Let $(h, O) \in \text{Aff } \mathcal{H}$. Suppose that f is not an O -coboundary and the μ -preserving transformation $G_{0, O}$ is mildly mixing. Then for each $\theta < \theta_{\text{diss}}$, the nonsingular Gaussian transformation $G_{\theta h, O}$ is weakly mixing, IDPFT, and of Krieger type III_1 . (Arano et al. 2021; Danilenko and Lemańczyk 2022).*

This theorem is generalized in a further work (Marrakchi and Vaes 2022) for some cases in which $G_{0, O}$ is weakly mixing but not mildly mixing. In these cases, $G_{0, O}$ has “mixing subsequences” along which the cocycle $(h^{(n)})_n$ is “proper.” In particular, an example of rigid weakly mixing Gaussian transformation $G_{0, O}$ is constructed such that the nonsingular Gaussian transformation $G_{h, O}$ is ergodic of type III_1 for some $h \in \mathcal{H}$.

Currently, the only known examples of ergodic nonsingular Gaussian transformations are either of type III_1 or II_1 . Therefore, though there is an obvious analogy between the theory of nonsingular Gaussian transformations and the theory of nonsingular Poisson suspensions, the latter theory looks more diverse. We illustrate this by the following theorem.

Theorem 9.13 *For each $K \in \{III_\lambda \mid 0 \leq \lambda \leq 1\} \sqcup \{II_\infty\}$, there is transformation $T \in \text{Aut}_2(Y, \nu)$ such that the nonsingular Poisson suspension T_* is 0-type, weakly mixing, and of Krieger type K , while the nonsingular Gaussian transformation $G_{\sqrt{\frac{d\nu \circ T}{d\nu}} - 1, U_T}$ is 0-type, weakly mixing, and of Krieger type III_1 . The unitary Koopman operators of the two nonsingular transformations are unitarily equivalent (Danilenko 2022).*

Spectral Theory for Nonsingular Systems

While the spectral theory for probability preserving systems is developed in depth, the spectral theory of nonsingular systems is still in its infancy. We discuss below some problems related to L^∞ -

spectrum which may be regarded as an analogue of the discrete spectrum. We also include results on computation of the maximal spectral type of the “nonsingular” Koopman operator for rank-one nonsingular transformations.

L^∞ - spectrum and Groups of Quasi-invariance

Let T be an ergodic nonsingular transformation of $(X, \mathcal{B}, \mu)$. A number $\lambda \in \mathbb{T}$ belongs to the L^∞ -spectrum $e(T)$ of T if there is a function $f \in L^\infty(X, \mu)$ with $f \circ T = \lambda f$. f is called an L^∞ -eigenfunction of T corresponding to λ . Denote by $\mathcal{E}(T)$ the group of all L^∞ -eigenfunctions of absolute value 1. It is a Polish group when endowed with the topology of converges in measure. If T is of type $I I_1$, then the L^∞ -eigenfunctions are $L^2(\mu')$ -eigenfunctions of T , where μ' is an equivalent invariant probability measure. Hence $e(T)$ is countable. Osikawa constructed in (Osikawa 1977) the first examples of ergodic nonsingular transformations with uncountable $e(T)$.

We state now a nonsingular version of the von Neumann-Halmos discrete spectrum theorem. Let $Q \subset \mathbb{T}$ be a countable infinite subgroup. Let K be a compact dual of Q_d , where Q_d denotes Q with the discrete topology. Let $k_0 \in K$ be the element defined by $k_0(q) = q$ for all $q \in Q$. Let $R : K \rightarrow K$ be defined by $Rk = k + k_0$. The system (K, R) is called a *compact group rotation*. The following theorem was proved in (Aaronson and Nadkarni 1987).

Theorem 10.1 Assume that the L^∞ -eigenfunctions of T generate the entire σ -algebra $\mathcal{B}$. Then T is isomorphic to a compact group rotation equipped with an ergodic quasi-invariant measure.

A natural question arises: Which subgroups of $\mathbb{T}$ can appear as $e(T)$ for an ergodic T ?

Theorem 10.2 $e(T)$ is a Borel subset of $\mathbb{T}$ and carries a unique Polish topology which is stronger than the usual topology on $\mathbb{T}$. The Borel structure of $e(T)$ under this topology agrees with the Borel structure inherited from $\mathbb{T}$. There is a Borel map $\psi : e(T) \ni \lambda \mapsto \psi_\lambda \in \mathcal{E}(T)$ such that

$\psi_\lambda \circ T = \lambda \psi_\lambda$ for each λ . Moreover, $e(T)$ is of Lebesgue measure 0 and it can have an arbitrary Hausdorff dimension (Aaronson 1997, 1983; Moore and Schmidt 1980).

A proper Borel subgroup E of $\mathbb{T}$ is called

- (i) *Weak Dirichlet* if $\limsup_{n \rightarrow \infty} \widehat{\lambda}(n) = 1$ for each finite complex measure λ supported on E
- (ii) *Saturated* if $\limsup_{n \rightarrow \infty} |\widehat{\lambda}(n)| \geq |\lambda(E)|$ for each finite complex measure λ on $\mathbb{T}$,

where $\widehat{\lambda}(n)$ denote the n -th Fourier coefficient of λ . Every countable subgroup of $\mathbb{T}$ is saturated.

Theorem 10.3 $e(T)$ is σ -compact in the usual topology on $\mathbb{T}$ (Host et al. 1991) and saturated (M  la 1983; Host et al. 1991).

It follows that $e(T)$ is weak Dirichlet (this fact was established earlier in (Schmidt 1982)).

It is not known if every Polish group continuously embedded in $\mathbb{T}$ as a σ -compact saturated group is the eigenvalue group of some ergodic nonsingular transformation. This is the case for the so-called H_2 -groups and the groups of quasi-invariance of measures on $\mathbb{T}$ (see below). Given a sequence n_j of positive integers and a sequence $a_j \geq 0$, the set of all $z \in \mathbb{T}$ such that $\sum_{j=1}^{\infty} a_j |1 - z^{n_j}|^2 < \infty$ is a group. It is called an H_2 -group. Every H_2 -group is Polish in an intrinsic topology stronger than the usual circle topology.

Theorem 10.4

- (i) Every H_2 -group is a saturated (and hence weak Dirichlet) σ -compact subset of $\mathbb{T}$ (Host et al. 1991).
- (ii) If $\sum_{j=0}^{\infty} a_j = +\infty$, then the corresponding H_2 -group is a proper subgroup of $\mathbb{T}$.
- (iii) If $\sum_{j=0}^{\infty} a_j (n_j/n_{j+1})^2 < \infty$, then the corresponding H_2 -group is uncountable.
- (iv) Any H_2 -group is $e(T)$ for an ergodic nonsingular compact group rotation T .

It is an open problem whether every eigenvalue group $e(T)$ is an H_2 -group. It is known however that $e(T)$ is close “to be an H_2 -group”: If a compact

subset $L \subset \mathbb{T}$ is disjoint from $e(T)$, then there is an H_2 -group containing $e(T)$ and disjoint from L .

Example 10.5 ((Aaronson and Nadkarni 1987), see also (Osikawa 1977)). Let (X, μ, T) be the nonsingular product odometer associated to a sequence $(2, v_j)_{j=1}^\infty$. Let n_j be a sequence of positive integers such that $n_j > \sum_{i < j} n_i$ for all j . For $x \in X$, we put $h(x) := n_{l(x)} - \sum_{j < l(x)} n_j$. Then h is a Borel map from X to the positive integers. Let S be the tower over T with height function h (see §3.3). Then $e(S)$ is the H_2 -group of all $z \in \mathbb{T}$ with $\sum_{j=1}^\infty v_j(0)v_j(1)|1 - z^{n_j}|^2 < \infty$.

It was later shown in (Host et al. 1991) that if $\sum_{j=1}^\infty v_j(0)v_j(1)(n_j/n_{j+1})^2 < \infty$, then the L^∞ -eigenfunctions of S generate the entire σ -algebra, i.e., S is isomorphic (measure theoretically) to a nonsingular compact group rotation.

Let μ be a finite measure on $\mathbb{T}$. Let $H(\mu) := \{z \in \mathbb{Z} \mid \delta_z * \mu \sim \mu\}$, where $*$ means the convolution of measures. Then H_μ is a group called the *group of quasi-invariance of μ* . It has a Polish topology whose Borel sets agree with the Borel sets which $H(\mu)$ inherits from $\mathbb{T}$, and the injection map of $H(\mu)$ into $\mathbb{T}$ is continuous. This topology is induced by the weak operator topology on the unitary group in the Hilbert space $L^2(\mathbb{T}, \mu)$ via the map $H(\mu) \ni z \mapsto U_z, (U_z f)(x) = \sqrt{(d(\delta_z * \mu)/d\mu)(x)} f(xz)$ for $f \in L^2(\mathbb{T}, \mu)$. Moreover, $H(\mu)$ is saturated (Host et al. 1991). If $\mu(H(\mu)) > 0$, then either $H(\mu)$ is countable or μ is equivalent to $\lambda_{\mathbb{T}}$ (Mandrekar and Nadkarni 1969).

Theorem 10.6 *Let μ be an ergodic with respect to the $H(\mu)$ -action by translations on $\mathbb{T}$. Then there is a compact group rotation (K, R) and a finite measure on K quasi-invariant and ergodic under R such that $e(R) = H(\mu)$. Moreover, there is a continuous one-to-one homomorphism $\psi : e(R) \rightarrow E(R)$ such that $\psi_\lambda \circ R = \lambda \psi_\lambda$ for all $\lambda \in e(R)$. (Aaronson and Nadkarni 1987).*

It was shown by Aaronson and Nadkarni (1987) that if $n_1 = 1$ and $n_j = a_j a_{j-1} \cdots a_1$ for positive integers $a_j \geq 2$ with $\sum_{j=1}^\infty a_j^{-1} < \infty$, then the transformation S from Example 10.5 does not

admit a continuous homomorphism $\psi : e(S) \rightarrow E(S)$ with $\psi_\lambda \circ T = \lambda \psi_\lambda$ for all $\lambda \in e(S)$. Hence $e(S) \neq H(\mu)$ for any measure μ satisfying the conditions of Theorem 10.6.

Assume that T is an ergodic nonsingular compact group rotation. Let $\mathcal{B}_0$ be the σ -algebra generated by a subcollection of eigenfunctions. Then $\mathcal{B}_0$ is invariant under T and hence a factor (see §12) of T . It is not known if every factor of T is of this form. It is not even known whether every factor of T must have nontrivial eigenvalues.

Koopman Unitary Operator for a Nonsingular System

Let $(X, \mathcal{B}, \mu, T)$ be a nonsingular dynamical system. In this subsection, we consider spectral properties of the Koopman operator U_T defined by (3). First, we note that the spectrum of T is the entire circle $\mathbb{T}$ (Nadkarni 1979). Next, if U_T has an eigenvector, then T is of type II_1 . Indeed, if there are $\lambda \in \mathbb{T}$ and $0 \neq f \in L^2(X, \mu)$ with $U_T f = \lambda f$, then the measure $\nu, d\nu(x) := |f(x)|^2 d\mu(x)$, is finite, T -invariant, and equivalent to μ . Hence if T is of type III or II_∞ , then the maximal spectral type σ_T of U_T is continuous. Another “restriction” on σ_T was found in (Roy 2009): No Foiaş-Strătilă measure is absolutely continuous with respect to σ_T if T is of type II_∞ . We recall that a symmetric measure σ on $\mathbb{T}$ possesses *Foiaş-Strătilă property* if for each ergodic probability-preserving system (Y, ν, S) and $f \in L^2(Y, \nu)$, if σ is the spectral measure of f , then f is a Gaussian random variable (Lemańczyk et al. 2000). For instance, measures supported on Kronecker sets possess this property.

As we have noted in §6, mixing (0-type) is an L^2 -spectral property for nonsingular transformations. Also, if T is infinite measure-preserving, then T is mixing if and only if $n^{-1} \sum_{i=0}^{n-1} U_T^{k_i} \rightarrow 0$ in the strong operator topology for each strictly increasing sequence $k_1 < k_2 < \cdots$ (Krengel and Sucheston 1969). This generalizes a well-known theorem of Blum and Hanson for probability-preserving maps. For comparison, we note that ergodicity is not an L^2 -spectral property of infinite measure-preserving systems.

Now let T be a rank-one nonsingular transformation associated with a sequence $(r_n, w_n, s_n)_{n=1}^\infty$ as in §3.4.

Theorem 10.7 *The spectral multiplicity of U_T is 1, and the maximal spectral type σ_T of U_T (up to a discrete measure in the case T is of type I_1) is the weak limit of the measures ρ_k defined as follows (Host et al. 1991; Choksi and Nadkarni 1994):*

$$d\rho_k(z) = \prod_{j=1}^k w_j(0) |P_j(z)|^2 dz,$$

where $P_j(z) := 1 + \sqrt{w_j(1)/w_j(0)} z^{-R_{1,j}} + \dots + \sqrt{w_j(m_j-1)/w_j(0)} z^{-R_{j-1,j}}$, $z \in \mathbb{T}$, $R_{i,j} := ih_{j-1} + s_j(0) + \dots + s_j(i)$, $1 \leq i \leq r_k - 1$ and h_j is the height of the j -th column.

Thus, the maximal spectral type of U_T is given by a so-called *generalized Riesz product*. We refer the reader to (Host et al. 1986, 1991; Choksi and Nadkarni 1994; Nadkarni 1998) for a detailed study of Riesz products: their convergence, mutual singularity, singularity to $\lambda_{\mathbb{T}}$, etc.

It was shown in (Aaronson and Nadkarni 1987) that $H(\sigma_T) \supset e(T)$ for any ergodic nonsingular transformation T . Moreover, σ_T is ergodic under the action of $e(T)$ by translations if T is isomorphic to an ergodic nonsingular compact group rotation. However it is not known:

- (i) Whether $H(\sigma_T) = e(T)$ for all ergodic T
- (ii) Whether ergodicity of σ_T under $e(T)$ implies that T is an ergodic compact group rotation

The first claim of Theorem 10.7 extends to the rank N nonsingular systems as follows: If T is an ergodic nonsingular transformation of rank N , then the spectral multiplicity of U_T is bounded by N (as in the finite measure-preserving case). It is not known whether this claim is true for a more general class of transformations which are defined as rank N but without the assumption that the Radon-Nikodym cocycle is constant on the tower levels.

Danilenko and Ryzhikov showed in (2010) that for each subset $E \subset \mathbb{N}$, there is an ergodic

conservative infinite measure-preserving transformation T such that the set of essential values of the multiplicity function of U_T is E . In a subsequent paper, Danilenko and Ryzhikov (2011) sharpened this result: For each subset $E \subset \mathbb{N} \cup \{\infty\}$, there is a mixing ergodic conservative infinite measure-preserving transformation T such that the set of essential values of the multiplicity function of U_T is E . We note that the analogous realization problem for spectral multiplicities of ergodic probability-preserving transformations is still open (Danilenko 2013).

In (Danilenko and Ryzhikov 2010), a mixing rank-one infinite measure-preserving transformation T was constructed such that the measures $\sigma_T, \sigma_T * \sigma_T, \sigma_T * \sigma_T * \sigma_T, \dots$ on $\mathbb{T}$ are mutually disjoint. Hence the unitary operator $U_T \oplus U_T^{\otimes 2} \oplus U_T^{\otimes 3} \oplus \dots$ has a simple spectrum.

El Abdalaoui and Nadkarni constructed an ergodic nonsingular transformation whose spectrum has Lebesgue component of multiplicity one (El Abdalaoui and Nadkarni 2016). The problem of existence of an ergodic nonsingular transformation with a simple Lebesgue spectrum is still open.

Koopman Unitary Operators Associated with Nonsingular Poisson Transformations

We recall a very important structural feature of Poisson spaces (see Neretin 1996): There is a canonical isometry between $L^2(\mu^*)$ and the symmetric Fock space $F(L^2(\mu))$ over $L^2(\mu)$. We recall that $F(L^2(\mu)) := \bigoplus_{n=0}^\infty L^2(\mu)^{\odot n}$, where $L^2(\mu)^{\odot 0} = \mathbb{C}$ and each factor $L^2(\mu)^{\odot n}$ is equipped with the normalized scalar product $n! \langle \cdot, \cdot \rangle_{L^2(\mu)^{\odot n}}$. There is an exponential map $\mathcal{E} : L^2(\mu) \rightarrow F(L^2(\mu))$, given by

$$\mathcal{E}(f) := \bigoplus_{n=0}^\infty \frac{1}{n!} f^{\odot n}.$$

The family $\{\mathcal{E}(f) \mid f \in L^2(\mu)\}$ is total in $F(L^2(\mu))$. Let $\text{Aff}_{\mathbb{R}}(L^2(\mu))$ be the subgroup of invertible affine operators in $L^2(\mu)$ that preserve invariantly the $\mathbb{R}$ -subspace $L^2_{\mathbb{R}}(\mu)$. Then $\text{Aff}_{\mathbb{R}}(L^2(\mu)) = L^2_{\mathbb{R}}(\mu) \rtimes \mathcal{U}_{\mathbb{R}}(L^2(\mu))$, where $\mathcal{U}_{\mathbb{R}}(L^2(\mu))$ is the group of unitary operators that preserve invariant $L^2_{\mathbb{R}}(\mu)$. The Weyl unitary

representation W of $\text{Aff}_{\mathbb{R}}(L^2(\mu))$ is well defined in $F(L^2(\mu))$ via the formula

$$W(f, V)\mathcal{E}(h) := e^{-\frac{1}{2}\|f\|_2 - \langle f, Vh \rangle} \mathcal{E}(f + Vh), \quad f \in L^2(\mu).$$

Given $T \in \text{Aut}_2(X, \mu)$, we have that $\left(\sqrt{\frac{d\mu \circ T}{d\mu}} - 1, U_T\right) \in \text{Aff}_{\mathbb{R}}(L^2(\mu))$.

Theorem 10.8 (Koopman operator associated with T_* (Danilenko et al. 2022a)). *If $T \in \text{Aut}_2(X, \mu)$, then under the canonical identification of $L^2(\mu^*)$ and $F(L^2(\mu))$,*

$$U_{T_*} = W\left(\sqrt{\frac{d\mu \circ T}{d\mu}} - 1, U_T\right).$$

Koopman Unitary Operators Associated with Nonsingular Gaussian Transformations

Let $\mathcal{H}$ be a separable infinite dimensional real Hilbert space. Denote by (X, μ) the probability space where the nonsingular Gaussian transformations $G_{h,O}$ for all $(h, O) \in \text{Aff } \mathcal{H}$ are defined (see §3.8). It is well known that there is a canonical isometry between $L^2(X, \mu)$ and the symmetric Fock space $F(\mathcal{H})$.

Theorem 10.9 (Koopman operator associated with $G_{h,O}$ (Danilenko and Lemańczyk 2022)). *If $(h, O) \in \text{Aff } \mathcal{H}$, then under the canonical identification of $L^2(X, \mu)$ and $F(\mathcal{H})$,*

$$U_{G_{h,O}} = W\left(\frac{1}{2}h, O\right).$$

It follows from Theorems 10.8 and 10.9 that each nonsingular Poisson transformation T_* is spectrally equivalent to the nonsingular Gaussian transformation $G\left(\sqrt{\frac{d\mu \circ T}{d\mu}} - 1\right)$.

Entropy and Other Invariants

Let T be an ergodic conservative nonsingular transformation of a standard probability space $(X, \mathcal{B}, \mu)$. If $\mathcal{P}$ is a finite partition of X , we define the entropy $H(\mathcal{P})$ of $\mathcal{P}$ as $H(\mathcal{P}) =$

$-\sum_{P \in \mathcal{P}} \mu(P) \log \mu(P)$. In the study of measure-preserving systems, the classical (Kolmogorov-Sinai) entropy proved to be a very useful invariant for isomorphism (Cornfeld et al. 1982). The key fact of the theory is that if $\mu \circ T = \mu$ then the limit $\lim_{n \rightarrow \infty} n^{-1} H\left(\bigvee_{i=1}^n T^{-i}\mathcal{P}\right)$ exists for every $\mathcal{P}$. However, if T does not preserve μ , the limit may no longer exist. Some efforts have been made to extend the use of entropy and similar invariants to the nonsingular domain. These include Krengel's entropy of conservative measure-preserving maps and its extension to nonsingular maps, Parry's entropy and Parry's nonsingular version of Shannon-McMillan-Breiman theorem, Poisson entropy, critical dimension by Mortiss and Dooley, etc. Unfortunately, these invariants are less informative than their classical counterparts and they are more difficult to compute.

Krengel and Parry's Entropies

Let S be a conservative measure-preserving transformation of a σ -finite measure space $(Y, \mathcal{E}, \nu)$. The *Krengel entropy* (Krengel 1967) of S is defined by

$$h_{\text{Kr}}(S) = \sup\{\nu(E)h(S_E) \mid 0 < \nu(E) < +\infty\},$$

where $h(S_E)$ is the Kolmogorov-Sinai entropy of S_E . It follows from Abramov's formula for the entropy of induced transformation that $h_{\text{Kr}}(S) = \mu(E)h(S_E)$ whenever E sweeps out, i.e., $\bigcup_{i \geq 0} S^{-i}E = X$. A generic transformation from $\text{Aut}_0(X, \mu)$ has entropy 0. Krengel raised a question in (Krengel 1967): Do there exist a zero entropy infinite measure-preserving S and a zero entropy finite measure-preserving R such that $h_{\text{Kr}}(S \times R) > 0$? This problem was solved in (Danilenko and Rudolph 2009) (a special case was announced by Silva and Thieullen in an October 1995 AMS conference (unpublished)):

- (i) If $h_{\text{Kr}}(S) = 0$ and R is distal, then $h_{\text{Kr}}(S \times R) = 0$;
- (ii) If R is not distal, then there is a rank-one transformation S with $h_{\text{Kr}}(S \times R) = \infty$.

We also note that if a conservative $S \in \text{Aut}_0(X, \mu)$ is squashable, i.e., it commutes

with another transformation R such that $v \circ R = cv$ for a constant $c \neq 1$, then $h_{\text{Kr}}(S)$ is either 0 or ∞ (Silva and Thieullen 1995).

Now let T be a type III ergodic transformation of $(X, \mathcal{B}, \mu)$. Silva and Thieullen define an entropy $h^*(T)$ of T by setting $h^*(T) := h_{\text{Kr}}(\tilde{T})$, where $\tilde{T}$ is the Maharam extension of T (see §5.2). Since $\tilde{T}$ commutes with transformations which “multiply” $\tilde{T}$ -invariant measure, it follows that $h^*(T)$ is either 0 or ∞ .

Let T be the standard III $_{\lambda}$ -odometer from Example 5.1 (i). Then $h^*(T) = 0$. The same is true for a so-called ternary product odometer associated with the sequence $(3, v_n)_{n=1}^{\infty}$, where $v_n(0) = v_n(2) = \lambda/(1 + 2\lambda)$ and $v_n(1) = \lambda/(1 + \lambda)$ (Silva and Thieullen 1995). It is not known however whether every ergodic nonsingular product odometer has zero entropy. On the other hand, it was shown in (Silva and Thieullen 1995) that $h^*(T) = \infty$ for every K -automorphism.

The Parry entropy (Parry 1969) of S is defined by

$$h_{\text{Pa}}(S) := \{H(S^{-1}\mathfrak{F}|\mathfrak{F}) | \mathfrak{F} \text{ is a } \sigma\text{-finite subalgebra of } \mathcal{B} \text{ such that } \mathfrak{F} \subset S^{-1}\mathfrak{F}\}.$$

Parry showed (1969) that $h_{\text{Pa}}(S) \leq h_{\text{Kr}}(S)$. It is still an open question whether the two entropies coincide. This is the case when S is of rank one (since $h_{\text{Kr}}(S) = 0$) and when S is quasi-finite (Parry 1969). The transformation S is called *quasi-finite* if there exists a subset of finite measure $A \subset Y$ such that the first return time partition $(A_n)_{n>0}$ of A has finite entropy. We recall that $x \in A_n \Leftrightarrow n$ is the smallest positive integer such that $T^n x \in A$. An example of non-quasi-finite ergodic infinite measure-preserving transformation was constructed in (Aaronson and Park 2009). A natural question is about existence of the maximal invariant σ -finite subalgebra of zero (Krengel or Parry) entropy. Such an algebra is called *the Krengel-Pinsker or the Parry-Pinsker factor* of T , respectively. Existence of the Krengel-Pinsker factors was proved in (Aaronson and Park 2009) for a special class of quasi-finite transformations called

LLB. This result was extended in (Janvresse et al. 2010) in the following way.

Theorem 11.1 *Let T be an ergodic quasi-finite transformation. Then either there is the Krengel-Pinsker factor of T which is also the Parry-Pinsker and the Poisson-Pinsker (see the next subsection below) factor of T or T is remotely infinite, i.e., there exists a sub- σ -algebra $\mathcal{F} \subset \mathcal{B}$ such that $T^{-1}\mathcal{F} \subset \mathcal{F}$, $\bigvee_{n>0} T^n \mathcal{F} = \mathcal{F}$ and the subalgebra $\bigwedge_{n>0} T^{-n} \mathcal{F}$ does not contain subsets of positive finite measure.*

Poisson Entropy

Poisson entropy for infinite measure-preserving transformations was introduced in (Roy 2005). Let (X, μ) be an infinite σ -finite space, and let T be a μ -preserving invertible transformation of X . The Poisson suspension T_* of T is well defined on a probability space (X^*, μ^*) and $\mu^* \circ T_* = \mu^*$ (see §3.7). It is ergodic if and only if T has no invariant sets of finite positive measure. It follows from Theorem 10.8 that U_{T_*} is the “exponent” of U_T . Hence, the maximal spectral type of U_{T_*} is $\sum_n \geq 0 (n!)^{-1} (\sigma_T)^{*n}$, where σ_T is a measure of the maximal spectral type of U_T .

Now the Poisson entropy $h_{\text{Po}}(T)$ of T is $h(T_*)$. The main question is: $h_{\text{Po}}(T)$ coincides with $h_{\text{Pa}}(T)$ or $h_{\text{Kr}}(T)$? It was shown in (Janvresse et al. 2010) that $h_{\text{Pa}}(T) \leq h_{\text{Po}}(T)$. If T is quasifinite or rank one, then the three entropies of T coincide (Janvresse et al. 2010). If T is the infinite Markov shift associated with a pair (P, π) for recurrent and irreducible P (see §3.6), then

$$\begin{aligned} h_{\text{Kr}}(T) &= h_{\text{Pa}}(T) \\ &= h_{\text{Po}}(T) = - \sum_{a \in A} \pi(a) \sum_{b \in A} P(a, b) \log P(a, b). \end{aligned}$$

If σ_T is singular or U_T has finite multiplicity, then $h_{\text{Po}}(T) = 0$ (Janvresse et al. 2010). It was also shown in (Janvresse et al. 2010) that given a nontrivial invariant σ -finite algebra $\mathcal{F}$ of $\mathcal{B}$, the natural $\mathcal{F}$ -relative version of Poisson entropy coincides with the relative (Krengel) entropy defined in (Danilenko and Rudolph 2009). Hence, if Krengel’s and the Poisson

entropies coincide on $T \upharpoonright \mathcal{F}$ for some $\mathcal{F}$, then $h_{\text{Kr}}(T) = h_{\text{Po}}(T)$. On the other hand, Janvresse and de la Rue constructed an ergodic conservative infinite measure-preserving transformation T such that $h_{\text{Kr}}(T) = 0$ but $h_{\text{Po}}(T) > 0$ (Janvresse and de la Rue 2012).

Definition 11.2 An ergodic measure-preserving transformation T of a σ -finite measure space $(X, \mathcal{B}, \mu)$ is said to have *totally positive Poisson entropy* if for each σ -finite T -invariant sub- σ -algebra $\mathcal{F} \subset \mathcal{B}$, the Poisson entropy of the system $(X, \mathcal{F}, \mu \upharpoonright \mathcal{F}, T)$ is strictly positive.

We note that the Poisson suspension of the system $(X, \mathcal{F}, \mu \upharpoonright \mathcal{F}, T)$ from the above definition is canonically a factor of $(\tilde{X}, \tilde{\mathcal{B}}, \tilde{\mu}, \tilde{T})$. Such factors of $\tilde{T}$ are called Poissonian. Roy showed in (Roy 2010) that if T has totally positive Poisson entropy, then T is of zero type.

Theorem 11.3 (Existence of the Poisson-Pinsker factor (Roy 2010)). *Let T be an ergodic measure-preserving transformation of an infinite σ -finite measure space $(X, \mathcal{B}, \mu)$. Then either T has totally positive entropy and $\tilde{T}$ is CPE or there is a σ -finite T -invariant sub- σ -algebra $\mathcal{E} \subset \mathcal{B}$ such that the Poisson suspension of $(X, \mathcal{F}, \mu \upharpoonright \mathcal{F}, T)$ is the Pinsker factor of $\tilde{T}$.*

If T has totally positive entropy, then the maximal spectral type of T is Lebesgue countable. If $h_{\text{Po}}(T) > 0$ and T possesses a Poisson-Pinsker factor, then the maximal spectral type of T in the orthocomplement to the Poisson-Pinsker factor is Lebesgue countable (Roy 2010).

Parry's Generalization of Shannon-MacMillan-Breiman Theorem

Let T be an ergodic transformation of a standard nonatomic probability space $(X, \mathcal{B}, \mu)$. Suppose that $f \circ T \in L^1(X, \mu)$ if and only if $f \in L^1(X, \mu)$. This means that there is $K > 0$ such that $K^{-1} < \frac{d\mu \circ T}{d\mu}(x) < K$ for a.a. x . Let $\mathcal{P}$ be a finite partition of X . Denote by $C_n(x)$ the atom of $\bigvee_{i=0}^n T^{-i}\mathcal{P}$ which

contains x . We put $\omega_{-1} = 0$. Parry shows in (Parry 1963) that

$$\begin{aligned} & \frac{\sum_{j=0}^n \log \mu(C_{n-j}(T^j x)) (\omega_j(x) - \omega_{j-1}(x))}{\sum_{i=0}^n \omega_i(x)} \\ & \rightarrow H\left(P \mid \bigvee_{i=1}^{\infty} T^{-i}\mathcal{P}\right) \\ & - \int_X \log E\left(\frac{d\mu \circ T}{d\mu} \mid \bigvee_{i=0}^{\infty} T^{-i}\mathcal{P}\right) d\mu \end{aligned}$$

for a.a. x . Parry also shows that under the aforementioned conditions on T ,

$$\begin{aligned} & \frac{1}{n} \left(\sum_{j=0}^n H\left(\bigvee_{i=0}^j T^{-i}\mathcal{P}\right) - \sum_{j=0}^{n-1} H\left(\bigvee_{i=1}^{j+1} T^{-i}\mathcal{P}\right) \right) \\ & \rightarrow H\left(\mathcal{P} \mid \bigvee_{i=1}^{\infty} T^{-i}\mathcal{P}\right) \end{aligned}$$

Critical Dimension

The critical dimension introduced by Mortiss (2003) measures the order of growth for sums of Radon-Nikodym derivatives. Let $(X, \mathcal{B}, \mu, T)$ be an ergodic nonsingular dynamical system. Given $\delta > 0$, let

$$X_{\delta} := \left\{ x \in X \mid \liminf_{n \rightarrow \infty} \frac{\sum_{i=0}^{n-1} \omega_i(x)}{n^{\delta}} > 0 \right\} \text{ and} \quad (6)$$

$$X^{\delta} := \left\{ x \in X \mid \liminf_{n \rightarrow \infty} \frac{\sum_{i=0}^{n-1} \omega_i(x)}{n^{\delta}} = 0 \right\}. \quad (7)$$

Then X_{δ} and X^{δ} are T -invariant subsets.

Definition 11.4 The *lower critical dimension* $\alpha(T)$ of T is $\sup\{\delta \mid \mu(X_{\delta}) = 1\}$. The *upper critical dimension* $\beta(T)$ of T is $\inf\{\delta \mid \mu(X^{\delta}) = 1\}$. (Mortiss 2003; Dooley and Mortiss 2009).

It was shown in (Dooley and Mortiss 2009) that the lower and upper critical dimensions are

invariants for isomorphism of nonsingular systems. Notice also that

$$\begin{aligned}\alpha(T) &= \liminf_{n \rightarrow \infty} \frac{\log(\sum_{i=1}^n \omega_i(x))}{\log n} \text{ and } \beta(T) \\ &= \limsup_{n \rightarrow \infty} \frac{\log(\sum_{i=1}^n \omega_i(x))}{\log n}.\end{aligned}$$

Moreover, $0 \leq \alpha(T) \leq \beta(T) \leq 1$. If T is of type II_1 , then $\alpha(T) = \beta(T) = 1$. If T is the standard III_λ -odometer from Example 5.1, then $\alpha(T) = \beta(T) = \log(1 + \lambda) - \frac{\lambda}{1+\lambda} \log \lambda$.

Theorem 11.5 (i) For every $\lambda \in [0, 1]$ and every $c \in [0, 1]$, there exists a nonsingular product odometer of type III_λ with critical dimension equal to c (Mortiss 2002).

(ii) For every $c \in [0, 1]$, there exists a nonsingular product odometer of type II_∞ with critical dimension equal to c (Dooley and Mortiss 2009).

Let T be the nonsingular product odometer associated with a sequence $(m_n, v_n)_{n=1}^\infty$. Let $s(n) = m_1 \cdots m_n$ and let $H(\mathcal{P}_n)$ denote the entropy of the partition of the first n coordinates with respect to μ . We now state a nonsingular version of Shannon-MacMillan-Breiman theorem for T from (Dooley and Mortiss 2009).

Theorem 11.6 Let m_i be bounded from above. Then

$$\begin{aligned}\text{(i)} \quad \alpha(T) &= \liminf_{n \rightarrow \infty} \inf \frac{-\sum_{i=1}^n \log m_i(x_i)}{\log s(n)} \\ &= \liminf_{n \rightarrow \infty} \frac{H(\mathcal{P}_n)}{\log s(n)} \text{ and} \\ \text{(ii)} \quad \beta(T) &= \limsup_{n \rightarrow \infty} \inf \frac{-\sum_{i=1}^n \log m_i(x_i)}{\log s(n)} \\ &= \limsup_{n \rightarrow \infty} \frac{H(\mathcal{P}_n)}{\log s(n)}\end{aligned}$$

for a.a. $x = (x_i)_{i \geq 1} \in X$.

It follows that in the case when $\alpha(T) = \beta(T)$, the critical dimension coincides with $\lim_{n \rightarrow \infty} \frac{H(\mathcal{P}_n)}{\log s(n)}$. In (Mortiss 2002), this expression (when it exists) was called *AC-entropy* (average

coordinate). It also follows from Theorem 11.6 that if T is a product odometer of bounded type, then $\alpha(T^{-1}) = \alpha(T)$ and $\beta(T^{-1}) = \beta(T)$. In (Dooley and Mortiss 2006), Theorem 11.6 was extended to a subclass of Markov odometers. Those results were further extended to so-called G -measures on product spaces (Mansfield and Dooley 2017) and a class of Bratteli-Vershik systems with multiple edges (Dooley and Hagihara 2012). The critical dimensions for nonsingular Bernoulli shifts (see §3.5) were investigated in (Dooley and Mortiss 2007):

Theorem 11.7 For any $\epsilon > 0$, there exists a nonsingular Bernoulli shift S from the Krengele class with $\alpha(S) < \epsilon$ and $\beta(S) > 1 - \epsilon$.

Nonsingular Restricted Orbit Equivalence

In (Mortiss 2000), Mortiss initiated study of a nonsingular version of Rudolph's restricted orbit equivalence (Rudolph 1985). This work is still in its early stages and does not yet deal with any form of entropy. However she introduced nonsingular orderings of orbits, defined sizes, and showed that much of the basic machinery still works in the nonsingular setting.

Nonsingular Joinings and Factors

The theory of joinings is a powerful tool to study probability-preserving systems and to construct striking counterexamples. It is interesting to study what part of this machinery can be extended to the nonsingular case. However, there are some principal obstacles for such extensions:

- There are too many quasi-invariant measures in view of the Glimm-Effros theorem (see Theorem 2.12).
- Ergodic components of a nonergodic joining need not be joinings of the original systems.

There are several ways to bypass these obstacles. The principal idea is to select always an appropriate (rather narrow) class of quasiinvariant measures under consideration or impose some

restrictions on the structure of joinings. This approach led to some progress in understanding twofold joinings and constructing prime systems of any Krieger type. As far as we know, the higher-fold nonsingular joinings have not been considered so far. It turned out however that an alternative coding technique, predating joinings in studying the centralizer and factors of the classical measure-preserving Chacón maps, can be used as well to classify factors of Cartesian products of some nonsingular Chacón maps.

Joinings, Nonsingular MSJ, and Simplicity

In this subsection, all measures are probability measures. A *nonsingular joining* of two nonsingular systems $(X_1, \mathcal{B}_1, \mu_1, T_1)$ and $(X_2, \mathcal{B}_2, \mu_2, T_2)$ is a measure $\hat{\mu}$ on the product $\mathcal{B}_1 \times \mathcal{B}_2$ that is nonsingular for $T_1 \times T_2$ and satisfies: $\hat{\mu}(A \times X_2) = \mu_1(A)$ and $\hat{\mu}(X_1 \times B) = \mu_2(B)$ for all $A \in \mathcal{B}_1$ and $B \in \mathcal{B}_2$. Clearly, the product $\mu_1 \times \mu_2$ is a nonsingular joining. Given a transformation $S \in C(T)$, the measure μ_S given by $\mu_S(A \times B) := \mu(A \cap S^{-1}B)$ is a nonsingular joining of (X, μ, T) and $(X, \mu \circ S^{-1}, T)$. It is called a *graph-joining* since it is supported on the graph of S . Another important kind of joinings that we are going to define now is related to factors of dynamical systems. Recall that given a nonsingular system $(X, \mathcal{B}, \mu, T)$, a sub- σ -algebra $\mathcal{A}$ of $\mathcal{B}$ such that $T^{-1}(\mathcal{A}) = \mathcal{A} \bmod \mu$ is called a *factor* of T . There is another, equivalent, definition. A nonsingular dynamical system $(Y, \mathcal{C}, \nu, S)$ is called a factor of T if there exists a measure-preserving map $\varphi : X \rightarrow Y$, called a *factor map*, with $\varphi T = S\varphi$ a.e. (If φ is only nonsingular, ν may be replaced with the equivalent measure $\mu \circ \varphi^{-1}$, for which φ is measure-preserving.) Indeed, the sub- σ -algebra $\varphi^{-1}(\mathcal{C}) \subset \mathcal{B}$ is T -invariant and, conversely, any T -invariant sub- σ -algebra of $\mathcal{B}$ defines a factor map by immanent properties of standard probability spaces; see (Aaronson 1997). If φ is a factor map as above, then μ has a disintegration with respect to φ , i.e., $\mu = \int \mu_y d\nu(y)$ for a measurable map $y \mapsto \mu_y$ from Y to the probability measures on X so that $\mu_y(\varphi^{-1}(y)) = 1$, the measure $\mu_{S\varphi(x)} \circ T$ is equivalent to $\mu_{\varphi(x)}$ and

$$\frac{d\mu \circ T}{d\mu}(x) = \frac{d\nu \circ S}{d\nu}(\varphi(x)) \frac{d\mu_{S\varphi(x)} \circ T}{d\mu_{\varphi(x)}}(x) \quad (8)$$

for a.e. $x \in X$. Define now the *relative product* $\hat{\mu} = \mu \times_{\varphi} \mu$ on $X \times X$ by setting $\hat{\mu} = \int \mu_y \times \mu_y d\nu(y)$. Then it is easy to deduce from (8) that $\hat{\mu}$ is a nonsingular self-joining of T .

We note however that the above definition of joining is too general to be satisfactory (as we noted in the introduction to this section). It does not reduce to the classical definition when we consider probability-preserving systems. Indeed, the following result was proved in (Rudolph and Silva 1989).

Theorem 12.1 *Let $(X_1, \mathcal{B}_1, \mu_1, T_1)$ and $(X_2, \mathcal{B}_2, \mu_2, T_2)$ be two finite measure-preserving systems such that $T_1 \times T_2$ is ergodic. Then for every λ , $0 < \lambda < 1$, there exists a nonsingular joining $\hat{\mu}$ of μ_1 and μ_2 such that $(T_1 \times T_2, \hat{\mu})$ is ergodic and of type III_{λ} .*

It is not known however if the nonsingular joining $\hat{\mu}$ can be chosen in every orbit equivalence class. In view of the above, Rudolph and Silva (1989) isolate an important subclass of joining. It is used in the definition of a nonsingular version of minimal self-joinings.

Definition 12.2 (i) A nonsingular joining $\hat{\mu}$ of (X_1, μ_1, T_1) and (X_2, μ_2, T_2) is *rational* if there exist measurable functions $c^1 : X_1 \rightarrow \mathbb{R}_+$ and $c^2 : X_2 \rightarrow \mathbb{R}_+$ such that

$$\begin{aligned} \hat{\omega}_1^{\mu}(x_1, x_2) &= \omega_1^{\mu_1}(x_1) \omega_1^{\mu_2}(x_2) c^1(x_1) \\ &= \omega_1^{\mu_1}(x_1) \omega_1^{\mu_2}(x_2) c^2(x_2) \quad \hat{\mu} \text{ a.e.} \end{aligned}$$

(ii) A nonsingular dynamical system $(X, \mathcal{B}, \mu, T)$ has *minimal self-joinings (MSJ)* over a class $\mathcal{M}$ of probability measures equivalent to μ , if for every $\mu_1, \mu_2 \in \mathcal{M}$, for every rational joining $\hat{\mu}$ of μ_1, μ_2 , a.e. ergodic component of $\hat{\mu}$ is either the product of its marginals or is the graph-joining supported on T^j for some $j \in \mathbb{Z}$.

Clearly, product measure, graph-joinings, and the relative products are all rational joinings. Moreover, a rational joining of finite measure-preserving systems is measure-preserving and a rational joining of type II_1 's is of type II_1 (Rudolph and Silva 1989). Thus we obtain the finite measure-preserving theory as a special case. As for the definition of MSJ, it depends on a class $\mathcal{M}$ of equivalent measures. In the finite measure-preserving case, $\mathcal{M} = \{\mu\}$. However, in the nonsingular case no particular measure is distinguished. We note also that Definition 12.2(ii) involves some restrictions on all rational joinings and not only ergodic ones as in the finite measure-preserving case. The reason is that an ergodic component of a nonsingular joining needs not be a joining of measures equivalent to the original ones (Aaronson 1987). For finite measure-preserving transformations, MSJ over $\{\mu\}$ is the same as the usual twofold MSJ (del Junco and Rudolph 1987).

A nonsingular transformation T on $(X, \mathcal{B}, \mu)$ is called *prime* if its only factors are $\mathcal{B}$ and $\{X, \emptyset\} \bmod \mu$. A (nonempty) class $\mathcal{M}$ of probability measures equivalent to μ is said to be *centralizer stable* if for each $S \in C(T)$ and $\mu_1 \in \mathcal{M}$, the measure $\mu_1 \circ S$ is in $\mathcal{M}$.

Theorem 12.3 *Let $(X, \mathcal{B}, \mu, T)$ be a ergodic nonatomic dynamical system such that T has MSJ over a class $\mathcal{M}$ that is centralizer stable. Then T is prime and the centralizer of T consists of the powers of T (Rudolph and Silva 1989).*

A question that arises is whether such a nonsingular dynamical system (not of type II_1) exist. Expanding on Ornstein's original construction from (Ornstein 1972), Rudolph and Silva construct in (Rudolph and Silva 1989), for each $0 \leq \lambda \leq 1$, a nonsingular rank-one transformation T_λ that is of type III_λ and that has MSJ over a class $\mathcal{M}$ that is centralizer stable. Type II_∞ examples with analogous properties were also constructed there. In this connection, it is worth to mention the example by Aaronson and Nadkarni (Aaronson and Nadkarni 1987) of II_∞ ergodic transformations that have no factor algebras on which the invariant measure is σ -finite (except for the entire

ones); however, these transformations are not prime.

A more general notion than MSJ, called *graph self-joinings* (GSJ), was introduced (Silva and Witte 1992): Just replace the words “on T^j for some $j \in \mathbb{Z}$ ” in Definition 12.2(ii) with “on S for some element $S \in C(T)$.” For finite measure-preserving transformations, GSJ over $\{\mu\}$ is the same as the usual twofold simplicity (del Junco and Rudolph 1987). The famous Veech theorem on factors of twofold simple maps (see del Junco and Rudolph 1987) was extended to nonsingular systems in (Silva and Witte 1992) as follows: If a system $(X, \mathcal{B}, \mu, T)$ has GSJ, then for every non-trivial factor $\mathcal{A}$ of T there exists a locally compact subgroup H in $C(T)$ (equipped with the weak topology) which acts smoothly (i.e., the partition into H -orbits is measurable) and such that $\mathcal{A} = \{B \in \mathcal{B} \mid \mu(hB\Delta B) = 0 \text{ for all } h \in H\}$. It follows that there is a cocycle φ from $(X, \mathcal{A}, \mu|_{\mathcal{A}})$ to H such that T is isomorphic to the φ -skew product extension $(T|_{\mathcal{A}})_\varphi$ (see §6.4). Of course, the ergodic nonsingular product odometers and, more generally, ergodic nonsingular compact group rotation (see § 10.1) have GSJ. However, except for this trivial case (the Cartesian square is non-ergodic) plus the systems with MSJ from (Rudolph and Silva 1989), no examples of type III systems with GSJ are known. In particular, no smooth examples have been constructed so far. This is in sharp contrast with the finite measure-preserving case where abundance of simple (or close to simple) systems are known (see del Junco and Rudolph 1987; Thouvenot 1995; Danilenko 2007).

Nonsingular Coding and Factors of Cartesian Products of Nonsingular Maps

As we have already noticed above, the nonsingular MSJ theory was developed in (Rudolph and Silva 1989) only for twofold self-joinings. The reasons for this were technical problems with extending the notion of rational joinings from twofold to n -fold self-joinings. However, while the twofold nonsingular MSJ or GSJ properties of T are sufficient to control the centralizer and the factors of T , it is not clear whether it

implies anything about the factors or centralizer of $T \times T$. Indeed, to control them one needs to know the fourfold joinings of T . However, even in the finite measure-preserving case, it is a long-standing open question whether twofold MSJ implies n -fold MSJ. That is why del Junco and Silva (2003) apply alternative – nonsingular coding – techniques to classify the factors of Cartesian products of nonsingular Chacón maps. The techniques were originally used in (del Junco 1978) to show that the classical Chacón map is prime and has trivial centralizer. They were extended to nonsingular systems in (del Junco and Silva 1995).

For each $0 < \lambda < 1$, we denote by T_λ the Chacón map (see §3.4) corresponding the sequence of probability vectors $w_n = (\lambda/(1 + 2\lambda), 1/(1 + 2\lambda), \lambda/(1 + 2\lambda))$ for all $n > 0$. One can verify that the maps T_λ are of type III_λ . (The classical Chacón map corresponds to $\lambda = 1$.) All of these transformations are defined on the same standard Borel space $(X, \mathcal{B})$. These transformations were shown to be power weakly mixing in (Adams et al. 2001). The centralizer of any finite Cartesian product of nonsingular Chacón maps is computed in the following theorem.

Theorem 12.4 *Let $0 < \lambda_1 < \dots < \lambda_k \leq 1$ and $n_1, \dots, n_k$ be positive integers. Then the centralizer of the Cartesian product $T_{\lambda_1}^{\otimes n_1} \times \dots \times T_{\lambda_k}^{\otimes n_k}$ is generated by maps of the form $U_1 \times \dots \times U_k$, where each U_i , acting on the n_i -dimensional product space X^{n_i} , is a Cartesian product of powers of T_{λ_i} or a coordinate permutation on X^{n_i} . (del Junco and Silva 2003).*

Let π denote the permutation on $X \times X$ defined by $\pi(x, y) = (y, x)$, and let $\mathcal{B}^{2\odot}$ denote the symmetric factor, i.e., $\mathcal{B}^{2\odot} = \{A \in \mathcal{B} \otimes \mathcal{B} \mid \pi(A) = A\}$. The following theorem classifies the factors of the Cartesian product of any two nonsingular type III_λ , $0 < \lambda < 1$, or type II_1 Chacón maps.

Theorem 12.5 *Let T_{λ_1} and T_{λ_2} be two nonsingular Chacón systems. Let $\mathcal{F}$ be a factor algebra of $T_{\lambda_1} \times T_{\lambda_2}$. (del Junco and Silva 2003).*

- (i) *If $\lambda_1 \neq \lambda$, then $\mathcal{F}$ is equal mod 0 to one of the four algebras $\mathcal{B} \otimes \mathcal{B}, \mathcal{B} \otimes \mathcal{N}, \mathcal{N} \otimes \mathcal{B}$, or $\mathcal{N} \otimes \mathcal{N}$, where $\mathcal{N} = \{\emptyset, X\}$.*
- (ii) *If $\lambda_1 = \lambda$, then $\mathcal{F}$ is equal mod 0 to one of the following algebras $\mathcal{B} \otimes \mathcal{B}, \mathcal{B} \otimes \mathcal{N}, \mathcal{N} \otimes \mathcal{B}, \mathcal{N} \otimes \mathcal{N}$, or $(T^m \times Id)\mathcal{B}^{2\odot}$ for some integer m .*

It is not hard to obtain type III_1 examples of Chacón maps for which the previous two theorems hold. However the construction of type II_∞ and type III_0 nonsingular Chacón transformations is more subtle as it needs the choice of ω_n to vary with n . In (Hamachi and Silva 2000), Hamachi and Silva construct type III_0 and type II_∞ examples; however, the only property proved for these maps is ergodicity of their Cartesian square. More recently, Danilenko (2004) has shown that all of them (in fact, a wider class of nonsingular Chacón maps of all types) are power weakly mixing.

In (Choksi et al. 1989), Choksi, Eigen, and Prasad asked whether there exists a zero entropy, finite measure-preserving mixing automorphism S , and a nonsingular type III automorphism T , such that $T \times S$ has no Bernoulli factors. Theorem 12.5 provides a partial answer (with a mildly mixing only instead of mixing) to this question: If S is the finite measure-preserving Chacón map and T is a nonsingular Chacón map as above, the factors of $T \times S$ are only the trivial ones, so $T \times S$ has no Bernoulli factors.

Joinings and MSJ for Infinite Measure-Preserving Systems

Adams, Friedman, and Silva introduced in (Adams et al. 1997) an infinite version of Chacón map T as a rank-one transformation associated with $(r_n, \omega_n, s_n)_{n=1}^\infty$ such that $r_n = 3$, $\omega_n(0) = \omega_n(1) = \omega_n(2)$, $s_n(0) = 0$, $s_n(1) = 1$ and $s_n(2) = 3h_n + 1$ for each $n > 0$. This is called the *infinite Chacón transformation*. Let (X, μ) be the space of T . Of course, $\mu(X) = \infty$. This transformation has infinite ergodic index (Adams et al. 1997), is not power weakly mixing and not multiply recurrent (Gruher et al. 2003), and has trivial centralizer (Janvresse et al. 2018). For each $d > 0$, Janvresse,

de la Rue, and Roy investigated $T^{\times d}$ -invariant measures on X^d which are *boundedly finite*. This means that for each d levels of every tower of the inductive construction, the measure of the Cartesian product of these levels is finite. The product $\otimes_{n=1}^d \mu$ and graph-joinings, i.e., measures of the form $(A_1, \dots, A_d) \mapsto \mu(S_1^{-1}A_1, \dots, S_d^{-1}A_d)$ for some transformations $S_1, \dots, S_d \in C(T)$, are boundedly finite. Moreover, T itself is uniquely ergodic in the sense that there is only one (up to scaling) boundedly finite T -invariant measure. It was shown in (Janvresse et al. 2018) that each ergodic $T^{\times d}$ -invariant boundedly finite measure is a direct product of so-called *diagonal measures*. Unlike the finite measure-preserving case, the class of diagonal measures does not reduce to the graphjoinings (with $S_1, \dots, S_d$ being the powers of T). It contains so-called *weird* measures whose marginals are singular to μ . As a corollary, it was proved that $C(T) = \{T^n \mid n \in \mathbb{Z}\}$ (Janvresse et al. 2018). Some of the weird measures are totally dissipative (supported on a single orbit), and some of them are conservative. Danilenko showed in (2018) that there is a conservative $T \times T$ -invariant boundedly finite measure with absolutely continuous marginals whose ergodic components are all weird. This phenomenon is impossible for another infinite version T of Chacon map constructed in (Janvresse et al. 2019). Its construction mimics the construction of the classical Chacon map so much that it gives a μ -conull subset X_∞ such that for each $d \geq 1$, each ergodic $T^{\times d}$ -invariant measure supported on X_∞^d is the direct product of several copies of μ and the graph-joinings generated by powers of T . As a corollary, we obtain that each boundedly finite d -fold self-joining of T (the marginals of a joining are absolutely continuous) is a convex combination of countably many ergodic joinings.

In (Danilenko 2018), the problems studied in (Janvresse et al. 2018) are considered from a different point of view. Let T be a homeomorphism of a locally compact Cantor space X . We assume that T is Radon uniquely ergodic, i.e., there is only one (up to scaling) Radon

T -invariant measure μ on X . A d -fold *Radon self-joining* of T is a Radon measure on X^d whose marginals (which may be nonsigma-finite) are equivalent to μ . We consider only Radon invariant measures and define Radon d -fold MSJ and Radon disjointness. Of course, each ergodic component of nonergodic Radon joining is Radon. However, it need not be a joining. Then the (C, F) -construction (see (Danilenko 2001a, 2007)) is used to produce a number of rank-one homeomorphisms of X whose ergodic joinings are explicitly described. The weird measures from (Janvresse et al. 2018) appear now as quasigraph Radon measures, i.e., they are graphs of equivariant maps whose domain and range are meager (and of zero measure) subsets of X . It is constructed an uncountable family of pairwise Radon disjoint infinite Chacon like Radon uniquely ergodic homeomorphisms with Radon MSJ. Moreover, every transformation of this family is Radon disjoint with its inverse (Danilenko 2018).

Smooth Nonsingular Transformations

Diffeomorphisms of smooth manifolds equipped with smooth measures are commonly considered as physically natural examples of dynamical systems. Therefore the construction of smooth models for various dynamical properties is a well-established problem of the modern (probability-preserving) ergodic theory. Unfortunately, the corresponding “nonsingular” counterpart of this problem is almost unexplored. We survey here several interesting facts related to the topic.

For $r \in \mathbb{N} \cup \{\infty\}$, denote by $\text{Diff}_+^r(\mathbb{T})$ the group of orientation-preserving C^r -diffeomorphisms of the circle $\mathbb{T}$. Endow this set with the natural Polish topology. Fix $T \in \text{Diff}_+^r(\mathbb{T})$. Since $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, there exists a C^1 -function $f: \mathbb{R} \rightarrow \mathbb{R}$ such that $T(x + \mathbb{Z}) = f(x) + \mathbb{Z}$ for all $x \in \mathbb{R}$. The *rotation number* $\rho(T)$

of T is the limit $\lim_{n \rightarrow \infty} \left(\underbrace{f \circ \dots \circ f}_{n \text{ times}} \right) (x) \pmod{1}$.

The limit exists and does not depend on the choice of x and f . It is obvious that T is nonsingular with respect to Lebesgue measure $\lambda_{\mathbb{T}}$. Moreover, if $T \in \text{Diff}_+^r(\mathbb{T})$ and $\rho(T)$ is irrational, then the dynamical system $(\mathbb{T}, \lambda_{\mathbb{T}}, T)$ is ergodic (Cornfeld et al. 1982). It is interesting to ask: which Krieger's type can such systems have?

Katznelson showed in (Katznelson 1977) that the subset of type III C^∞ -diffeomorphisms and the subset of type II C^∞ -diffeomorphisms are dense in $\text{Diff}_+^\infty(\mathbb{T})$. Hawkins and Schmidt refined the idea of Katznelson from (Katznelson 1977) to construct, for every irrational number $\alpha \in [0, 1]$ which is not of constant type (i.e., in whose continued fraction expansion the denominators are not bounded) a transformation $T \in \text{Diff}_+^2(\mathbb{T})$ which is of type III₁ and $\rho(T) = \alpha$ (Hawkins and Schmidt 1982). It should be mentioned that class C^2 in the construction is essential, since it follows from a remarkable result of Herman that if $T \in \text{Diff}_+^3(\mathbb{T})$, then under some condition on α (which determines a set of full Lebesgue measure), T is measure theoretically (and topologically) conjugate to a rotation by $\rho(T)$ (Herman 1979b). Hence T is of type II₁.

In (Hawkins 1983), Hawkins shows that every smooth paracompact manifold of dimension ≥ 3 admits a type III _{λ} diffeomorphism for every $\lambda \in [0, 1]$. This extends a result of Herman (Herman 1979a) on the existence of type III₁ diffeomorphisms in the same circumstances.

It is also of interest to ask: Which free ergodic flows are associated with smooth dynamical systems of type III₀? Hawkins proved that any free ergodic C^∞ -flow on a smooth, connected, paracompact manifold is the associated flow for a C^∞ -diffeomorphism on another manifold (of higher dimension) (Hawkins 1990a).

A nice result was obtained in (Katznelson 1979; Hawkins and Woods 1984). If $T \in \text{Diff}_+^2(\mathbb{T})$ and the rotation number of T has unbounded continued fraction coefficients, then $(\mathbb{T}, \lambda_{\mathbb{T}}, T)$ is ITPFI. Moreover, a converse also holds: Given a nonsingular product odometer R , the set of orientation-preserving C^∞ -diffeomorphisms of the circle which are orbit equivalent to R is C^∞ -dense in the Polish set of

all C^∞ -orientation-preserving diffeomorphisms with irrational rotation numbers. In contrast to that, Hawkins constructs in (Hawkins 1982) a type III₀ C^∞ -diffeomorphism of the 4-dimensional torus which is not ITPFI.

Examples of n -to-1 conservative ergodic nonsingular C^∞ -endomorphisms on the 2-torus, not admitting an equivalent σ -finite invariant measure, were constructed in (Hawkins and Silva 1991). In (Avila and Bochi 2007), it is shown that a C^1 generic expanding map of $\mathbb{T}$ has no absolutely continuous σ -finite invariant measure.

Kosloff in (2021) showed that $\mathbb{T}^2$ admits a C^1 Anosov diffeomorphism of type III₁ with respect to Lebesgue measure. We recall that this phenomenon is impossible in the class of conservative $C^{1+\alpha}$ Anosov diffeomorphisms because by a theorem of Gurevich and Oseledets, every such transformation is of type II₁ (with respect to Lebesgue measure). In a later work (Kosloff 2018), he extended this result to $\mathbb{T}^d$ for every $d > 3$. The case $d = 3$ remains open.

Miscellaneous Topics

Let T be an ergodic measure-preserving transformation of an infinite σ -finite nonatomic measure space $(X, \mathcal{B}, \mu)$.

On Normalizing Constants for Ergodic Theorem

Replacing μ with an equivalent probability measure, one can deduce from the Hurewicz ergodic theorem that the average $\frac{1}{n} \sum_{i=0}^{n-1} f(T^i x)$ converges to 0 a.e. for each function $f \in L^1(X, \mu)$. In view of that, a natural question arises: Is there a sequence of positive numbers $(a_n)_{n=1}^\infty$ such that

$$\frac{1}{a_n} \sum_{i=0}^{n-1} f(T^i x) \rightarrow \int f d\mu \text{ a.e.} \quad (9)$$

for each $f \in L^1(X, \mu)$? Aaronson answered this question negatively in (Aaronson 1977b). He showed that if there is a sequence of positive numbers $(a_n)_{n=1}^\infty$ and a single integrable function $f \geq 0$ with $\int f d\mu > 0$ such that (9) holds, then $\mu(X) < \infty$.

Thus, no normalizing constants in the ergodic theorem for infinite measure-preserving transformations exist. Since then, other forms of convergence of ergodic averages for a given sequence have been studied, for which the reader may refer to (Aaronson 1981, 1997; Aaronson and Zweimüller 2014; Thaler 1998; Thaler and Zweimüller 2006) and the references therein. The situation is somewhat different in the case of symmetric Birkhoff sums, i.e., when we replace the average $\sum_{i=0}^{n-1}$ in (9) with $\sum_{|i|<n}$. It is shown in (Aaronson et al. 2017) that though (9) with $\sum_{|i|<n}$ instead of $\sum_{i=0}^{n-1}$ does not hold for all $f \in L^1(X, \mu)$, there are examples of ergodic conservative infinite measure-preserving transformations $(X, \mathfrak{B}, \mu, T)$ admitting sequences $(a_n)_n$ with $a_n \rightarrow +\infty$ and such that

$$\limsup_n \frac{1}{2} a_n \sum_{|i|<n} f(T^i x) = \int f d\mu \quad \text{and} \\ \liminf_n \frac{1}{2} a_n \sum_{|i|<n} f(T^i x) = \frac{1}{2} \int f d\mu \quad a.e.$$

for each $f \in L^1_+(X, \mu)$. See also (Kosloff 2017).

Around King's Weak Closure Theorem

We recall that if S is probability preserving rank-one map, then $C(S)$ is the weak closure of the set $\{S^n \mid n \in \mathbb{Z}\}$ (King theorem, (King 1986)). It is still unclear whether this theorem extends to the infinite measure-preserving rank-one transformations. However, there are some classes of infinite rank-one maps for which it is true: zero-type maps and partially bounded maps.

Definition 14.1 A σ -finite self-joining (of order 2) of T is a σ -finite $T \times T$ -invariant measure λ on $(X \times X, \mathfrak{B} \otimes \mathfrak{B})$ such that $\lambda(A \times X) = \lambda(X \times A) = \lambda(A)$ for all $A \in \mathfrak{B}$ of finite measure. If for each ergodic σ -finite self-joining λ of T , there is $n \in \mathbb{Z}$ such that $\lambda(A \times B) = \mu(A \times T^{-n}B)$, then T is said to have *minimal σ -finite self-joinings (of order 2)*.

The above concept of MSJ permits to control the μ -preserving centralizer $C_0(T)$ of T . If T has MSJ, then $C_0(T) = \{T^n \mid n \in \mathbb{Z}\}$. It was shown by Ryzhikov and Thouvenot (Ryzhikov and Thouvenot 2015) that each zero-type

transformation of rank one has σ -finite MSJ. Since the rank-one transformations are non-squashable, it follows that the centralizer of each zero-type rank-one transformation is just its powers.

Definition 14.2 Let T be a rank-one transformation associated with $(r_n, \omega_n, s_n)_{n=1}^\infty$. It is called *partially bounded* if there is $L > 0$ such that $r_n \leq L, \omega_n(0) = \dots = \omega_n(r_n - 1)$, $\max_{0 \leq i < j < r_n - 1} |s_n(i) - s_n(j)| < L, s_n(r_n - 1) = 0$ and $\min_{0 \leq i < r_n - 1} s_n(i) \geq h_n$ for each $n > 0$ (Gaebler et al. 2018).

It was shown in (Gaebler et al. 2018) that for each partially bounded transformation, the centralizer consists of just the powers. Of course, the family of partially bounded transformations does not intersect the set of zero-type rank-one maps.

Asymmetry and Bergelson's Question

We say that T is *asymmetric* if T is not isomorphic to T^{-1} . Explicit examples of asymmetric infinite rank-one transformations are constructed in (Danilenko and Ryzhikov 2012; Ryzhikov 2014) (see there for asymmetric maps which embed into a flow). It was shown in (Gaebler et al. 2018) that if T is a partially bounded rank-one transformation, then T is isomorphic to T^{-1} if and only if $s_n(i) = s_n(r_n - 2 - i)$ for all $i = 0, \dots, r_n - 2$ eventually in n .

Bergelson asked: Is there T of infinite ergodic index such that $T \times T^{-1}$ is not ergodic? Of course, such a T is asymmetric. The question is still open. However some partial progress was achieved in (Clancy et al. 2016; Danilenko 2016a). In (Clancy et al. 2016), an example of a rank-one T was constructed such that $T \times T$ is ergodic but $T \times T^{-1}$ is not. Similar examples appeared also in (Danilenko 2016a). However, they do not answer Bergelson's question because T has ergodic index 2 in these examples. It was also shown in (Danilenko 2016a) that within the class of infinite Markov shifts, the answer on Bergelson's question is negative. As for the rank-one transformations, it was shown in (Clancy et al. 2016) that $T \times T^{-1}$ is always conservative.

Ergodicity of Powers

Let T be a rank-one infinite measure-preserving transformation associated with $(r_n, \omega_n, s_n)_{n=1}^\infty$. (Hence $\omega_n(0) = \dots = \omega_n(r_n - 1)$.) For each $n > 0$, we denote by C_n the set of bottom levels of all copies of $(n - 1)$ -th tower in the n -th tower. Thus, formally, $C_n := \{0\} \cup \{jh_{n-1} + \sum_{i=0}^{j-1} s_n(i) | j = 1, \dots, r_n - 1\}$. The following was proved in (Danilenko 2019b).

Theorem 14.3 (Ergodicity of powers of rank-one transformations).

- (i) If T^d is ergodic, then for each divisor p of d there are infinitely many n such that some $c \in C_n$ is not divisible by p .
- (ii) If $(r_n)_{n=1}^\infty$ is bounded and for each divisor p of a positive integer d , there are infinitely many n such that p does not divide some $c \in C_n$, then T^d is ergodic.
- (iii) If the sequence $(r_n)_{n=1}^\infty$ is bounded, then T is totally ergodic if and only if for each $d > 1$, there are infinitely many $n > 0$ such that some element $c \in C_n$ is not divisible by d .

Rigidity Sequences

Adams proved in (2015) that if S is an ergodic probability-preserving transformation such that $S^{n_k} \rightarrow \text{Id}$ weakly for some sequence $(n_k)_{k=1}^\infty$ of positive integers, then there exists a *power rationally weakly mixing* infinite measure-preserving transformation T such that $T^{n_k} \rightarrow \text{Id}$ weakly. The converse assertion is obvious – just pass to the Poisson suspension. (An infinite measure-preserving T is called *power rationally weakly mixing* if for each finite sequence of nonzero integers $l_1, \dots, l_k$, the product transformation $T^{l_1} \times \dots \times T^{l_k}$ is rationally weakly mixing.) Thus the class of rigidity sequences for the ergodic probability-preserving transformations equals to the class of rigidity sequences for the ergodic infinite measure-preserving ones.

Directional Recurrence

Given an ergodic infinite measure-preserving $\mathbb{Z}^d$ -action $T = (T_g)_{g \in \mathbb{Z}^d}$, it seems natural to study the dynamics of individual transformations T_g , $g \in \mathbb{Z}^d$. For instance, one of the natural questions

is to describe the set of those $g \in G$ such that T_g is recurrent. Since T_g is recurrent if and only if T_{ng} is recurrent for every $n \in \mathbb{Z} \setminus \{0\}$, it makes sense to talk about *recurrent directions* in $\mathbb{Z}^d$ or *rational recurrent directions* in the group $\mathbb{R}^d$ containing $\mathbb{Z}^d$ as a standard lattice. Following Milnor's general idea of directional dynamics (Milnor 1988), Johnson and Şahin introduced in (2015) a concept of directional recurrence of T along an arbitrary (including irrational) direction in $\mathbb{R}^d$. They showed that the set $R(T)$ of recurrent directions is a G_δ -subset, produced examples of T with trivial and nontrivial $R(T)$, and asked about a description of all possible $R(T)$ when T runs through the ergodic $\mathbb{Z}^d$ -action. Some partial answers were obtained in (Danilenko 2017): Given a G_δ -subset Δ of the real projective space $P(\mathbb{R}^d)$ and a countable subset $D \subset \Delta$, there is a rank-one action T with $D \subset R(T) \subset \Delta$. However, in general, the problem remains open.

Applications. Connections with Other Fields

In this – final – section, we shed light on numerous mathematical sources of nonsingular systems. They come from the theory of stochastic processes, random walks, locally compact Cantor systems, horocycle flows on hyperbolic surfaces, von Neumann algebras, statistical mechanics, representation theory for groups and anti-commutation relations, etc. We also note that nonsingular systems can appear in the context of probability-preserving dynamics, for example, in §11.1 some criteria for distality are given in terms of the Krengel entropy. See also §15.1 and §15.3 below.

Mild Mixing

An ergodic finite measure-preserving dynamical system $(X, \mathcal{B}, \mu, T)$ is called *mildly mixing* if for each nontrivial T -invariant σ -algebra $\mathcal{A} \subset \mathcal{B}$, the restriction $T|_{\mathcal{A}}$ is not rigid. For equivalent definitions and extensions to actions of locally compact groups, we refer to (Aaronson 1997) and (Schmidt and Walters 1982). There is an interesting criterion of the mild mixing that involves nonsingular

systems: T is mildly mixing if and only if for each ergodic nonsingular transformation S , the product $T \times S$ is ergodic (Furstenberg and Weiss 1978). Furthermore, T is mildly mixing if and only if for each ergodic nonsingular transformation S , the product $T \times S$ is orbit equivalent to S (Hawkins and Silva 1997/98). In particular, the associated flows of S and $T \times S$ are isomorphic. Moreover, if R is a nonsingular transformation such that $R \times S$ is ergodic for any ergodic nonsingular S , then R is of type II_1 (and mildly mixing) (Schmidt and Walters 1982). In this context, we note that for every ergodic infinite measure-preserving transformation T there is an ergodic Markov shift S such that $T \times S$ is not conservative, hence not ergodic (Aaronson et al. 1979); also that S can be chosen to be rank-one and rigid (Elyze et al. 2018).

Ergodicity of Gaussian Cocycles

Let T be an ergodic (equivalently, weakly mixing) measure-preserving Gaussian transformation on a standard probability space $(X, \mathfrak{B}, \mu)$, and let H be the corresponding invariant Gaussian subspace of the real Hilbert space $L_0^2(X, \mu) := L^2(X, \mu) \ominus \mathbb{R}$. The following conjecture was stated in (Lemańczyk et al. 2001): For each function $f \in H$, either f is a T -coboundary (equivalently, a Gaussian coboundary, i.e., the transfer function belongs to H), or the skew product transformation T_f acting on the product space $(X \times \mathbb{R}, \mu \times \text{Leb})$ is ergodic. We note that T_f preserves the infinite measure $\mu \times \text{Leb}$. The affirmative answer was obtained in (Danilenko and Lemańczyk 2022) (and independently in (Marrakchi and Vaes 2022)) under an assumption that T is mildly mixing. Some examples of ergodic T_f with rigid weakly mixing T were constructed in (Marrakchi and Vaes 2022). However, in the general setting, the problem remains open.

Disjointness and Furstenberg's Class $\mathcal{W}^\perp$

Two probability-preserving systems (X, μ, T) and (Y, ν, S) are called *disjoint* if $\mu \times \nu$ is the only $T \times S$ -invariant probability measure on $X \times Y$ whose coordinate projections are μ and ν , respectively.

Furstenberg in (1967) initiated studying the class $\mathcal{W}^\perp$ of transformations disjoint from all weakly mixing ones. Let $\mathcal{D}$ denote the class of distal transformations and $\mathcal{M}(\mathcal{W}^\perp)$ the class of multipliers of $\mathcal{W}^\perp$ (for the definitions, see (Glasner 1994)). Then $\mathcal{D} \subset \mathcal{M}(\mathcal{W}^\perp) \subset \mathcal{W}^\perp$. In (Danilenko and Lemańczyk 2005; Lemańczyk and Parreau 2003), it was shown by constructing explicit examples that these inclusions are strict. We record this fact here because nonsingular ergodic theory was the key ingredient of the arguments in the two papers pertaining to the theory of probability-preserving systems. The examples are of the form $T_{\varphi, S}(x, y) = (Tx, S_{\varphi(x)}y)$, where T is an ergodic rotation on (X, μ) , $(S_g)_{g \in G}$ a mildly mixing action of a locally compact group G on Y and $\varphi : X \rightarrow G$ a measurable map. Let W_φ denote the Mackey action of G associated with φ , and let (Z, κ) be the space of this action. The key observation is that there exists an affine isomorphism of the simplex of $T_{\varphi, S}$ -invariant probability measures whose pullback on X is μ and the simplex of $W_\varphi \times S$ quasi-invariant probability measures whose pullback on Z is κ and whose Radon-Nikodym cocycle is measurable with respect to Z . This is a far-reaching generalization of Furstenberg theorem on relative unique ergodicity of ergodic compact group extensions.

Symmetric Stable and Infinitely Divisible Stationary Processes

Rosinsky in (1995) established a remarkable connection between structural studies of stationary stochastic processes and ergodic theory of nonsingular transformations (and flows). For simplicity, we consider only real processes in discrete time. Let $X = (X_n)_{n \in \mathbb{Z}}$ be a measurable stationary symmetric α -stable (SaS) process, $0 < \alpha < 2$. This means that any linear combination $\sum_{k=1}^n a_k X_{j_k}, j_k \in \mathbb{Z}, a_k \in \mathbb{R}$ has an SaS-distribution. (The case $\alpha = 2$ corresponds to Gaussian processes.) Then the process admits a spectral representation

$$X_n = \int_Y f_n(y) M(dy), n \in \mathbb{Z}, \quad (10)$$

where $f_n \in L^\alpha(Y, \mu)$ for a standard σ -finite measure space $(Y, \mathcal{B}, \mu)$ and M is an independently scattered random measure on $\mathcal{B}$ such that $E \exp(iuM(A)) = \exp(-|u|^\alpha \mu(A))$ for every $A \in \mathcal{B}$ of finite measure. By (Rosinsky 1995), one can choose the kernel $(f_n)_{n \in \mathbb{Z}}$ in a special way: There are a μ -nonsingular transformation T and measurable maps $\varphi : X \rightarrow \{-1, 1\}$ and $f \in L^\alpha(Y, \mu)$ such that $f_n = U^n f$, $n \in \mathbb{Z}$, where U is the isometry of $L^\alpha(X, \mu)$ given by $Ug = \varphi \cdot (d\mu \circ T/d\mu)^{1/\alpha} \cdot g \circ T$. If, in addition, the smallest T -invariant σ -algebra containing $f^{-1}(\mathcal{B}_{\mathbb{R}})$ coincides with $\mathcal{B}$ and $\text{Supp}\{f \circ T^n : n \in \mathbb{Z}\} = Y$, then the pair (T, φ) is called minimal. It turns out that minimal pairs always exist. Moreover, two minimal pairs (T, φ) and (T', φ') representing the same SaS process are equivalent in some natural sense (Rosinsky 1995). Then one can relate ergodic-theoretical properties of (T, φ) to probabilistic properties of $(X_n)_{n \in \mathbb{Z}}$. For instance, let $Y = C \sqcup D$ be the Hopf decomposition of Y (see Theorem 2.2). We let $X_n^D := \int_D f_n(y) M(dy)$ and $X_n^C := \int_C f_n(y) M(dy)$. Then we obtain a unique (in distribution) decomposition of X into the sum $X^D + X^C$ of two independent stationary SaS-processes.

Another kind of decomposition was considered in (Samorodnitsky 2005). Let P be the largest invariant subset of Y such that $T|_P$ has a finite invariant measure. Partitioning Y into P and $N := Y \setminus P$ and restricting the integration in (10) to P and N , we obtain a unique (in distribution) decomposition of X into the sum $X^P + X^N$ of two independent stationary SaS-processes. Notice that the process X is ergodic if and only if $\mu(P) = 0$.

Roy considered a more general class of *infinitely divisible (ID)* stationary processes (Roy 2007). Using Maruyama's representation of the characteristic function of an ID process X without Gaussian part, he singled out the Lévy measure Q of X . Then Q is a shift invariant σ -finite measure on $\mathbb{R}^{\mathbb{Z}}$. Decomposing the dynamical system $(\mathbb{R}^{\mathbb{Z}}, \tau, Q)$ in various natural ways (Hopf decomposition, 0-type and positive type, so-called "rigidity free" part, and its complement), he obtains corresponding decompositions for the process X . Here τ stands for the shift on $\mathbb{R}^{\mathbb{Z}}$.

Poisson Suspensions of Infinite Measure-Preserving Transformations

Poisson suspensions over infinite measure-preserving transformations (see (Roy 2005, 2009)) are widely used in statistical mechanics to model ideal gas, Lorentz gas, etc. (see (Cornfeld et al. 1982)). Being a particular case of the Poisson suspensions over nonsingular transformations (see §3.7), they are always probability preserving. Together with the Gaussian dynamical systems, they are also an important source of examples and counterexamples in ergodic theory. Due to a close similarity with the well-studied Gaussian systems, a natural question arises: Are there ergodic Poisson suspensions whose ergodic self-joinings are Poisson? Such suspensions are called PAP. They are analogue of GAG in the theory of Gaussian systems (Lemańczyk et al. 2000). Janvresse, de la Rue, and Roy constructed PAP suspension in (Janvresse et al. 2017) (see also (Parreau and Roy 2008)). The example of an infinite measure-preserving T with "minimal self-joinings" from (Janvresse et al. 2019) plays a crucial role in their construction. We also mention a result of Meyerovitch (Meyerovitch 2013) related to weak mixing of infinite measure-preserving systems and Poisson suspensions: If T is a conservative ergodic infinite measure-preserving transformation, then the direct product of T with the Poisson suspension T_* of T is ergodic.

Recurrence of Random Walks with Nonstationary Increments

Using nonsingular ergodic theory, one can introduce the notion of recurrence for random walks obtained from certain nonstationary processes. Let T be an ergodic nonsingular transformation of a standard probability space $(X, \mathcal{B}, \mu)$, and let $f : X \rightarrow \mathbb{R}^n$ be a measurable function. Define for $m \geq 1$, $Y_m : X \rightarrow \mathbb{R}^n$ by $Y_m := \sum_{n=0}^{m-1} f \circ T^n$. In other words, $(Y_m)_{m \geq 1}$ is the random walk associated with the (nonstationary) process $(f \circ T^n)_{n \geq 0}$. Let us call this random walk *recurrent* if the cocycle f of T is recurrent (see §5.5). It was shown in (Schmidt 1984) that in the case $\mu \circ T = \mu$, i.e., the process is stationary, this definition is equivalent to the standard one.

Boundaries of Random Walks

Boundaries of random walks on groups retain valuable information on the underlying groups (amenability, entropy, etc.) and enable one to obtain integral representation for harmonic functions of the random walk (Zimmer 1977, 1978; Kaimanovich and Vershik 1983). Let G be a locally compact group and ν a probability measure on G . Let T denote the (one-sided) shift on the probability space $(X, \mathcal{B}_X, \mu) := (G, \mathcal{B}_G, \nu)^{\mathbb{Z}_+}$ and $\varphi : X \rightarrow G$ a measurable map defined by $(y_0, y_1, \dots) \mapsto y_0$. Let T_φ be the φ -skew product extension of T acting on the space $(X \times G, \mu \times \lambda_G)$ (for noninvertible transformations, the skew product extension is defined in the very same way as for invertible ones; see §5.5). Then T_φ is isomorphic to the *homogeneous random walk* on G with jump probability ν . Let $\mathcal{I}(T_\varphi)$ denote the sub- σ -algebra of T_φ -invariant sets, and let $\mathcal{F}(T_\varphi) := \bigcap_{n>0} T_\varphi^{-n}(\mathcal{B}_X \otimes \mathcal{B}_G)$. The former is called the *Poisson boundary* of T_φ and the latter one is called the *tail boundary* of T_φ . Notice that a nonsingular action of G by inverted right translations along the second coordinate is well defined on each of the two boundaries. The two boundaries (or, more precisely, the G -actions on them) are ergodic. The Poisson boundary is the Mackey range of φ (as a cocycle of T). Hence the Poisson boundary is amenable (Zimmer 1978). If the support of ν generates a dense subgroup of G , then the corresponding Poisson boundary is weakly mixing (Aaronson and Lemańczyk 2005). As for the tail boundary, we first note that it can be defined for a wider family of *nonhomogeneous* random walks. This means that the jump probability ν is no longer fixed and a sequence $(\nu_n)_{n>0}$ of probability measures on G is considered instead. Now let $(X, \mathcal{B}_X, \mu) := \prod_{n>0} (G, \mathcal{B}_G, \nu_n)$. The one-sided shift on X may not be nonsingular now. Instead of it, we consider the tail equivalence relation $\mathcal{R}$ on X and a cocycle $\alpha : \mathcal{R} \rightarrow G$ given by $\alpha(x, y) = x_1 \cdots x_n y_n^{-1} \cdots y_1$, where $x = (x_i)_{i>0}$ and $y = (y_i)_{i>0}$ are $\mathcal{R}$ -equivalent and n is the smallest integer such that $x_i = y_i$ for all $i > n$. The tail boundary of the random walk on G with time-dependent jump probabilities $(\nu_n)_{n>0}$ is the Mackey G -action associated with α . In the case

of homogeneous random walks, this definition is equivalent to the initial one. Connes and Woods showed (Connes and Woods 1989) that the tail boundary is always amenable and AT. It is unknown whether the converse holds for general G . However, it is true for $G = \mathbb{R}$ and $G = \mathbb{Z}$: the class of AT-flows coincides with the class of tail boundaries of the random walks on $\mathbb{R}$, and a similar statement holds for $\mathbb{Z}$ (Connes and Woods 1989). Jaworski showed (Jaworski 1994) that if G is countable and a random walk is homogeneous, then the tail boundary of the random walk possesses a so-called SAT-property (which is stronger than AT).

Stationary Actions

Let $T = (T_g)_{g \in G}$ be a continuous action of a countable group G on a compact metrizable space X . By Markov-Kakutani theorem, if G is amenable, then there is an invariant probability Borel measure on X . If G is nonamenable, such a measure does not necessarily exist. However, if κ is a probability measure on G whose support generates G as a semigroup, then there is always a T -quasiinvariant probability measure μ on X such that

$$\sum_{g \in G} \kappa(g) \frac{d\mu \circ T_g^{-1}}{d\mu}(x) = 1 \text{ for a.e. } x \in X.$$

μ is called a κ -stationary measure. For a deep theory of stationary actions and its applications, we refer to (Furman 2002; Furstenberg and Glasner 2010) and references therein. However, if G is Abelian or, more generally, nilpotent, then each κ -stationary action is invariant under T . Thus, there are no stationary $\mathbb{Z}$ -actions except for the probability-preserving ones. That is why we do not discuss them in this survey.

Classifying σ -finite Ergodic Invariant Measures

The description of ergodic finite invariant measures for topological (or, more generally, standard Borel) systems is a well-established problem in the classical ergodic theory (Cornfeld et al. 1982). On the other hand, it seems impossible to obtain any useful information about the system by

analyzing the set of all ergodic quasi-invariant (or just σ -finite invariant) measures because this set is wildly huge (see §2.6). The situation changes if we impose some restrictions on the measures. For instance, if the system under question is a homeomorphism (or a topological flow) defined on a locally compact Polish space, then it is natural to consider the class of (σ -finite) invariant Radon measures, i.e., measures taking finite values on the compact subsets. We give two examples.

First, the seminal results of Giordano, Putnam, and Skau on the topological orbit equivalence of compact Cantor minimal systems were extended to locally compact Cantor minimal (l.c.c.m.) systems in (Danilenko 2001b; Matui 2002). Given a l.c.c.m. system X , we denote by $\mathcal{M}(X)$ and $\mathcal{M}_1(X)$ the set of invariant Radon measures and the set of invariant probability measures on X . Notice that $\mathcal{M}_1(X)$ may be empty (Danilenko 2001b). It was shown in (Matui 2002) that two systems X and X' are topologically orbit equivalent if and only if there is a homeomorphism of X onto X' which maps bijectively $\mathcal{M}(X)$ onto $\mathcal{M}(X')$ and $\mathcal{M}_1(X)$ onto $\mathcal{M}_1(X')$. Thus $\mathcal{M}(X)$ retains an important information on the system – it is “responsible” for the topological orbit equivalence of the underlying systems. Uniquely ergodic l.c.c.m. systems (with unique up to scaling infinite invariant Radon measure) were constructed in (Danilenko 2001b).

The second example is related to study of the smooth horocycle flows on tangent bundles of hyperbolic surfaces. Let $\mathbb{D}$ be the open disk equipped with the hyperbolic metric $|dz|/(1 - |z|^2)$, and let $\text{Möb}(\mathbb{D})$ denote the group of Möbius transformations of $\mathbb{D}$. A hyperbolic surface can be written in the form $M := \Gamma \backslash \text{Möb}(\mathbb{D})$, where Γ is a torsion-free discrete subgroup of $\text{Möb}(\mathbb{D})$. Suppose that Γ is a nontrivial normal subgroup of a lattice Γ_0 in $\text{Möb}(\mathbb{D})$. Then M is a regular cover of the finite volume surface $M_0 := \Gamma_0 \backslash \text{Möb}(\mathbb{D})$. The group of deck transformations $G = \Gamma_0/\Gamma$ is finitely generated. The horocycle flow $(h_t)_{t \in \mathbb{R}}$ and the geodesic flow $(g_t)_{t \in \mathbb{R}}$ defined on the unit tangent bundle $T^1(\mathbb{D})$ descend naturally to flows, say h and g , on $T^1(M)$. We consider the problem of classification of the h -invariant Radon measures

on M . According to Ratner, h has no finite invariant measures on M if G is infinite (except for measures supported on closed orbits). However there are infinite invariant Radon measures, for instance, the volume measure. In the case when G is free Abelian and Γ_0 is cocompact, every homomorphism $\varphi : G \rightarrow \mathbb{R}$ determines a unique up to scaling ergodic invariant Radon measure (e.i.r.m.) m on $T^1(M)$ such that $m \circ dD = \exp(\varphi(D))m$ for all $D \in G$ (Babillot and Ledrappier 1998) and every e.i.r.m. arises this way (Sarig 2004). Moreover, all these measures are quasi-invariant under g . In the general case, an interesting bijection is established in (Ledrappier and Sarig 2007) between the e.i.r.m. which are quasi-invariant under g and the “non-trivial minimal” positive eigenfunctions of the hyperbolic Laplacian on M . This result was extended in the recent work (Landesberg and Lindenstrauss 2022).

Von Neumann Algebras

There is a fascinating and productive interplay between nonsingular ergodic theory and von Neumann algebras. The two theories alternately influenced development of each other. Let $(X, \mathcal{B}, \mu, T)$ be a nonsingular dynamical system. Given $\varphi \in L^\infty(X, \mu)$ and $j \in \mathbb{Z}$, we define operators A_φ and U_j on the Hilbert space $L^2(Z \times \mathbb{Z}, \mu \times \nu)$ by setting

$$(A_\varphi f)(x, i) := \varphi(T^i x) f(x, i), \quad (U_j f)(x, i) := f(x, i - j)$$

Then $U_j A_\varphi U_j^* = A^{\varphi \circ T^j}$. Denote by $\mathcal{M}$ the von Neumann algebra (i.e., the weak closure of the $*$ -algebra) generated by A_φ , $\varphi \in L^\infty(X, \mu)$ and U_j , $j \in \mathbb{Z}$. If T is ergodic and aperiodic, then $\mathcal{M}$ is a factor, i.e., $\mathcal{M} \cap \mathcal{M}' = \mathbb{C}1$, where $\mathcal{M}'$ denotes the algebra of bounded operators commuting with $\mathcal{M}$. It is called a *Krieger's factor*. Murray-von Neumann-Connes' type of $\mathcal{M}$ is exactly the Krieger's type of T . The flow of weights of $\mathcal{M}$ is isomorphic to the associated flow of T . Two Krieger's factors are isomorphic if and only if the underlying dynamical systems are orbit equivalent (Krieger 1976b). Moreover, a number of important problems in the theory of

von Neumann algebras such as classification of subfactors, computation of the flow of weights and Connes' invariants, outer conjugacy for automorphisms, etc. are intimately related to the corresponding problems in nonsingular orbit theory. We refer to (Moore 1982; Feldman and Moore 1977; Giordano and Skandalis 1985a, b; Hamachi and Kosaki 1993; Danilenko and Hamachi 2000) for details.

Representations of CAR

Representations of canonical anticommutation relations (CAR) is one of the most elegant and useful chapters of mathematical physics, providing a natural language for many body quantum physics and quantum field theory. By a representation of CAR, we mean a sequence of bounded linear operators $a_1, a_2, \dots$ in a separable Hilbert space $\mathcal{H}$ such that $a_j a_k + a_k a_j = 0$ and $a_j a_k^* + a_k^* a_j = \delta_{j,k}$.

Consider $\{0, 1\}$ as a group with addition mod 2. Then $X = \{0, 1\}^{\mathbb{N}}$ is a compact Abelian group. Let $\Gamma := \{x = (x_1, x_2, \dots) : \lim_{n \rightarrow \infty} x_n = 0\}$. Then Γ is a dense countable subgroup of X . It is generated by elements γ_k whose k -coordinate is 1 and the other ones are 0. Γ acts on X by translations. Let μ be an ergodic Γ -quasi-invariant measure on X . Let $(C_k)_{k \geq 1}$ be Borel maps from X to the group of unitary operators in a Hilbert space $\mathcal{H}$ satisfying $C_k^*(x) = C_k(x + \delta_k)$, $C_k(x) C_l(x + \delta_l) = C_l(x) C_k(x + \delta_k)$, $k \neq l$ for a.a. x . In other words, $(C_k)_{k \geq 1}$ defines a cocycle of the Γ -action. We now put $\mathcal{H} := L^2(X, \mu) \otimes \mathcal{H}$ and define operators a_k in $\mathcal{H}$ by setting

$$(a_k f)(x) = (-1)^{x_1 + \dots + x_{k-1}} (1 - x_k) C_k(x) \\ \times \sqrt{\frac{d\mu \circ \delta_k}{d\mu}}(x) f(x + \delta_k),$$

where $f: X \rightarrow \mathcal{H}$ is an element of $\widetilde{\mathcal{H}}$ and $x = (x_1, x_2, \dots) \in X$. It is easy to verify that $a_1, a_2, \dots$ is a representation of CAR. The converse was established in (Gårding and Wightman 1954; Golodets 1969): Every factor-representation (this means that the von Neumann algebra generated by all a_k is a factor) of CAR can be represented as above for some ergodic measure μ , Hilbert space

$\mathcal{H}$, and a Γ -cocycle $(C_k)_{k \geq 1}$. Moreover, using nonsingular ergodic theory Golodets (Golodets 1969) constructed for each $k = 2, 3, \dots, \infty$, an irreducible representation of CAR such that $\dim \mathcal{H} = k$. This answered a question of Gårding and Wightman (Gårding and Wightman 1954) who considered only the case $k = 1$.

Unitary Representations of Locally Compact Groups

Nonsingular actions appear in a systematic way in the theory of unitary representations of groups. Let G be a locally compact second countable group and H a closed normal subgroup of G . Suppose that H is commutative (or, more generally, of type I, see (Dixmier 1969)). Then the natural action of G by conjugation on H induces a Borel G -action, say α , on the dual space $\widehat{H}$ – the set of unitarily equivalent classes of irreducible unitary representations of H . If now $U = (U_g)_{g \in G}$ is a unitary representation of G in a separable Hilbert space, then by applying Stone decomposition theorem to $U \upharpoonright H$ one can deduce that α is nonsingular with respect to a measure μ of the “maximal spectral type” for $U \upharpoonright H$ on $\widehat{H}$. Moreover, if U is irreducible, then α is ergodic. Whenever μ is fixed, we obtain a one-to-one correspondence between the set of cohomology classes of irreducible cocycles for α with values in the unitary group on a Hilbert space $\mathcal{H}$ and the subset of $\widehat{G}$ consisting of classes of those unitary representations V for which the measure associated to $V \upharpoonright H$ is equivalent to μ . This correspondence is used in both directions. From information about cocycles, we can deduce facts about representations and vice versa (Kirillov 1978; Dixmier 1969).

Further Directions

While some of the results that we have cited for nonsingular $\mathbb{Z}$ -actions extend to actions of locally compact Polish groups (or subclasses of Abelian or amenable ones), many natural questions remain open in the general setting. For instance, what is the Rokhlin lemma, or the pointwise ergodic theorem (for some obstacles toward extension of the

ratio ergodic theorem to nonsingular actions of arbitrary amenable groups, see (Hochman 2013); a weak version of this theorem was proved recently in (Danilenko 2019a)), or the definition of entropy for nonsingular actions of general countable amenable groups? The theory of abstract nonsingular equivalence relations (Feldman and Moore 1977) or, more generally, nonsingular groupoids (Ramsay 1971) and polymorphisms (Vershik 1983) is also a beautiful part of nonsingular ergodic theory that has nice applications: description of semifinite traces of AF-algebras, classification of factor representations of the infinite symmetric group (Vershik and Kerov 1985), path groups (Albeverio et al. 1983), etc. Nonsingular ergodic theory needs different tools for the most part when we pass from $\mathbb{Z}$ -actions to noninvertible endomorphisms or, more generally, semigroup actions. Several concrete open problems are mentioned throughout the entry.

Bibliography

- Aaronson J (1977a) Rational ergodicity and a metric invariant for Markov shifts. *Israel J Math* 27:93–123
- Aaronson J (1977b) On the ergodic theory of non-integrable functions and infinite measure spaces. *Israel J Math* 27(2):163–173
- Aaronson J (1979) Rational ergodicity, bounded rational ergodicity and some continuous measures on the circle. *Israel J Math* 33:181–197
- Aaronson J (1981) The asymptotic distributional behaviour of transformations preserving infinite measures. *J Analyse Math* 39:203–234
- Aaronson J (1983) The eigenvalues of nonsingular transformations. *Israel J Math* 45:297–312
- Aaronson J (1987) The intrinsic normalizing constants of transformations preserving infinite measures. *J. Analyse Math.* 49:239–270
- Aaronson J (1997) An introduction to infinite ergodic theory. American Mathematical Society, Providence, RI, USA
- Aaronson J (2013) Rational weak mixing in infinite measure spaces. *Ergodic Theory Dyn Syst* 33:1611–1643
- Aaronson J (2016) Conditions for rational weak mixing. *Stoch Dyn* 16:1660004
- Aaronson J, Lemańczyk M (2005) Exactness of Rokhlin endomorphisms and weak mixing of Poisson boundaries, algebraic and topological dynamics. In: *Contemporary mathematics*, 385. American Mathematical Society, Providence, pp 77–88
- Aaronson J, Nadkarni M (1987) L_∞ eigenvalues and L_2 spectra of nonsingular transformations. *Proc Lond Math Soc* 55(3):538–570
- Aaronson J, Nakada H (2000) Multiple recurrence of Markov shifts and other infinite measure-preserving transformations. *Israel J Math* 117:285–310
- Aaronson J, Park KK (2009) Predictability, entropy and information of infinite transformations, *Fundam Math* 206:1–21
- Aaronson J, Weiss B (2004) On Herman's theorem for ergodic, amenable group extensions of endomorphisms. *Ergodic Theory Dyn Syst* 24(5):1283–1293
- Aaronson J, Weiss B (2018) Distributional limits of positive, ergodic stationary processes and infinite ergodic transformations. *Ann Inst Henri Poincaré Probab Stat* 54:879–906
- Aaronson J, Zweimüller R (2014) Limit theory for some positive stationary processes with infinite mean, English, with English and French summaries. *Ann Inst Henri Poincaré Probab Stat* 50(1):256–284
- Aaronson J, Lin M, Weiss B (1979) Mixing properties of Markov operators and ergodic transformations, and ergodicity of Cartesian products. *Israel J Math* 33:198–224
- Aaronson J, Kosloff Z, Weiss B (2017) Symmetric Birkhoff sums in infinite ergodic theory. *Ergodic Theory Dyn Syst* 37:2394–2416
- Abdalaoui EH el, Nadkarni MG (2016) A non-singular transformation whose spectrum has Lebesgue component of multiplicity one. *Ergodic Theory Dyn Syst* 36:671–681
- Adams TM (1998) Smorodinsky's conjecture on rank-one mixing. *Proc Am Math Soc* 126:739–744
- Adams TM (2015) Rigidity sequences of power rationally weakly mixing transformations, preprint, arXiv:1503.05806v1
- Adams T, Silva CE (2015) On infinite transformations with maximal control of ergodic two-fold product powers. *Israel J Math* 209:929–948
- Adams T, Silva CE (2016) Weak rational ergodicity does not imply rational ergodicity. *Israel J Math* 214:491–506
- Adams T, Silva CE (2018) Weak mixing for infinite measure invertible transformations. In: *Ergodic theory and dynamical systems in their interactions with arithmetics and combinatorics*, Lecture notes in Mathematics, 2213, Springer, Cham, pp 327–349
- Adams S, Elliott GA, Giordano T (1994) Amenable actions of groups. *Trans Am Math Soc* 344:803–822
- Adams T, Friedman N, Silva CE (1997) Rank-one weak mixing for nonsingular transformations. *Israel J Math* 102:269–281
- Adams T, Friedman N, Silva CE (2001) Rank one power weak mixing for nonsingular transformations. *Ergodic Theory Dyn Syst* 21:1321–1332
- Ageev ON, Silva CE (2001) Genericity of rigid and multiply recurrent infinite measure-preserving and nonsingular transformations, *Proceedings of the 16th*

- summer conference on general topology and its applications. *Topology Proc* 26(2):357–365, 2001/02
- Albeverio S, Hoegh-Krohn R, Testard D, Vershik AM (1983) Factorial representations of Path groups. *J Funct Anal* 51:115–231
- Alpern S, Prasad VS (1990) Return times for nonsingular measurable transformations. *J Math Anal Appl* 152: 470–487
- Araki H, Woods EJ (1968) A classification of factors. *Pub RIMS, Ser A* 3:51–130
- Arano Y, Isono Y, Marrakchi M (2021) Ergodic theory of affine isometric actions on Hilbert spaces. *Geom Funct Anal* 31(5):1013–1094
- Atkinson G (1976) Recurrence of co-cycles and random walks. *J Lond Math Soc* 13:486–488
- Avila A, Bochi J (2007) Generic expanding maps without absolutely continuous invariant σ -finite measure. *Math Res Lett* 14(5):721–730
- Avraham-Re'em N (2022) On absolutely continuous invariant measures and Krieger-type of Markov subshifts. *J d'Anal Math* 147:201–253
- Babillot M, Ledrappier F (1998) Geodesic paths and horocycle flow on abelian covers. In: *Lie groups and ergodic theory* (Mumbai, 1996), Tata Institute of Fundamental Research studies in Mathematics, 14, Tata Institute of Fundamental Research, Bombay, pp 1–32
- Berendschot T, Vaes S (2022a) Nonsingular Bernoulli actions of arbitrary Krieger type. *Anal PDE* 15:1313–1373
- Berendschot T, Vaes S (2022b) Bernoulli actions of type III_0 with prescribed associated flow. *Journal of the Institute of Mathematics of Jussieu*, to appear
- Bergelson V, Leibman A (1996) Polynomial extensions of van der Waerden's and Semerédi's theorems. *J Am Math Soc* 9:725–753
- Bezuglyi SI, Golodets VY (1985) Groups of measure space transformations and invariants of outer conjugation for automorphisms from normalizers of type III full groups. *J Funct Anal* 60(3):341–369
- Bezuglyi SI, Golodets VY (1991) Weak equivalence and the structures of cocycles of an ergodic automorphism. *Publ Res Inst Math Sci* 27(4):577–625
- Björklund M, Kosloff Z, Vaes S (2021) Ergodicity and type of nonsingular Bernoulli actions. *Invent Math* 224(2): 573–625
- Bonanno C, Giulietti P, Lenci M (2018) Infinite mixing for one-dimensional maps with an indifferent fixed point. *Nonlinearity* 31:5180–5213
- Bowles A, Fidkowski L, Marinello A, Silva CE (2001) Double ergodicity of nonsingular transformations and infinite measure-preserving staircase transformations. *Ill J Math* 45(3):999–1019
- Bozgan F, Sanchez A, Silva CE, Stevens D, Wang J (2015) Subsequence bounded rational ergodicity of rank-one transformations. *Dynamical Syst* 30:70–84
- Bruin H, Melbourne I, Terhesiu D (2019) Rates of mixing for non-Markov infinite measure semi-flows. *Trans Am Math Soc* 371:7343–7386
- Chacon RV, Friedman NA (1965) Approximation and invariant measures. *Z Wahrscheinlichkeitstheorie und Verw Gebiete* 3:286–295
- Choksi JR, Kakutani S (1979) Residuality of ergodic measurable transformations and of ergodic transformations which preserve an infinite measure. *Ind Univ Math J* 28:453–469
- Choksi JR, Nadkarni MG (1994) The maximal spectral type of a rank one transformation. *Can Math Bull* 37(1):29–36
- Choksi JR, Nadkarni MG (2000) Genericity of nonsingular transformations with infinite ergodic index. *Colloq Math* 84/85:195–201
- Choksi JR, Prasad VS (1983) Approximation and Baire category theorems in ergodic theory. In: *Measure theory and its applications* (Sherbrooke, Que., 1982), Lecture notes in mathematics 1033. Springer, Berlin, pp 94–113
- Choksi JR, Hawkins JM, Prasad VS (1987) Abelian cocycles for nonsingular ergodic transformations and the genericity of type III_1 transformations. *Monat fur Math* 103:187–205
- Choksi J, Eigen S, Prasad V (1989) Ergodic theory on homogeneous measure algebras revisited, Measure and measurable dynamics (Rochester, NY, 1987). In: *Contemporary Mathematics* 94. American Mathematical Society, Providence, pp 73–85
- Clancy J, Friedberg R, Kasmalkar I, Loh I, Padurariu T, Silva CE, Vasudevan S (2016) Ergodicity and conservativity of products of infinite transformations and their inverses. *Colloq Math* 143:271–291
- Connes A (1975) On the hierarchy of W. Krieger. *Ill J Math* 19:428–432
- Connes A, Krieger W (1977) Measure space automorphisms, the normalizers of their full groups, and approximate finiteness. *J Funct Anal* 24(4):336–352
- Connes A, Woods EJ (1985) Approximately transitive flows and ITPFI factors. *Ergodic Theory Dyn Syst* 5(2):203–236
- Connes A, Woods EJ (1989) Hyperfinite von Neumann algebras and Poisson boundaries of time dependent random walks. *Pacific J Math* 37:225–243
- Connes A, Feldman J, Weiss B (1981) An amenable equivalence relation is generated by a single transformation. *Ergodic Theory Dyn Syst* 1:431–450
- Cornfeld IP, Fomin SV, Sinai YG (1982) *Ergodic theory*, Grundlehren der Mathematischen Wissenschaften, 245. Springer-Verlag, New York
- Dai I, Garcia X, Padurariu T, Silva CE (2015) On rationally ergodic and rationally weakly mixing rank-one transformations. *J Ergodic Theory Dyn Syst* 35:1141–1164
- Danilenko AI (1995) The topological structure of Polish groups and groupoids of measure space transformations. *Publ Res Inst Math Sci* 31(5):913–940
- Danilenko AI (1998) Quasinormal subrelations of ergodic equivalence relations. *Proc Am Math Soc* 126(11): 3361–3370
- Danilenko AI (2001a) Funny rank one weak mixing for nonsingular Abelian actions. *Israel J Math* 121:29–54
- Danilenko AI (2001b) Strong orbit equivalence of locally compact Cantor minimal systems. *Int J Math* 12:113–123
- Danilenko AI (2004) Infinite rank one actions and nonsingular Chacon transformations. *Ill J Math* 48(3): 769–786

- Danilenko AI (2007) (C, F)-actions in ergodic theory, in “Geometry and Dynamics of Groups and Spaces”. *Progr Math* 265:325–351
- Danilenko AI (2013) A survey on spectral multiplicities of ergodic actions. *Ergodic Theory Dyn Syst* 33: 81–117
- Danilenko AI (2016a) Finite ergodic index and asymmetry for infinite measure preserving actions. *Proc Am Math Soc* 144:2521–2532
- Danilenko AI (2016b) Actions of finite rank: weak rational ergodicity and partial rigidity. *J Ergodic Theory Dyn Syst* 36:2138–2171
- Danilenko AI (2017) Directional recurrence and directional rigidity for infinite measure preserving actions of nilpotent lattices. *Ergodic Theory Dyn Syst* 37: 1841–1861
- Danilenko AI (2018) Infinite measure preserving transformations with Radon MSJ. *Israel J Math* 228:21–51
- Danilenko AI (2019a) Weak mixing for nonsingular Bernoulli actions of countable amenable groups. *Proc Am Math Soc* 147:4439–4450
- Danilenko AI (2019b) Rank-one actions, their (C,F)-models and constructions with bounded parameters. *J d’Anal Math* 139(2):697–749
- Danilenko AI (2021) On the bounded cohomology for ergodic nonsingular actions of amenable groups. *Israel J Math* 243(1):421–436
- Danilenko AI (2022) Krieger’s type for ergodic nonsingular Poisson actions of non-(T) locally compact groups. *Ergodic Theory Dyn Systems*, to appear
- Danilenko AI, del Junco A (2011) Almost continuous orbit equivalence for non-singular homeomorphisms. *Israel J Math* 183:165–188
- Danilenko AI, Golodets VY (1996) On extension of cocycles to normalizer elements, outer conjugacy, and related problems. *Trans Am Math Soc* 348(12): 4857–4882
- Danilenko AI, Hamachi T (2000) On measure theoretical analogues of the Takesaki structure theorem for type III factors. *Colloq Math* 84/85:485–493
- Danilenko AI, Kosloff Z (2022) Krieger’s type of nonsingular Poisson suspensions and IDPFT systems. *Proc Am Math Soc* 150(4):1541–1557
- Danilenko AI, Lemańczyk M (2005) A class of multipliers for $\mathcal{W}^\perp$. *Isr J Undergrad Math* 148:137–168
- Danilenko AI, Lemańczyk M (2019) K-property for Maharam extensions of non-singular Bernoulli and Markov shifts. *Ergodic Theory Dyn Syst* 39(12): 3292–3321
- Danilenko AI, Lemańczyk M (2022) Ergodic cocycles of IDPFT systems and non-singular Gaussian actions. *Ergodic Theory Dyn Syst* 42(5):1624–1654
- Danilenko AI, Park KK (2011) Rank-one flows of transformations with infinite ergodic index. *Proc Am Math Soc* 139:201–207
- Danilenko AI, Rudolph DJ (2009) Conditional entropy theory in infinite measure and a question of Krengel. *Israel J Math* 172:93–117
- Danilenko AI, Ryzhikov VV (2010) Spectral multiplicities of infinite measure preserving transformations. *Funct Anal Appl* 44:161–170
- Danilenko AI, Ryzhikov VV (2011) Mixing constructions with infinite invariant measure and spectral multiplicities. *Ergodic Theory Dyn Syst* 31:853–873
- Danilenko AI, Ryzhikov VV (2012) On self-similarities of ergodic flows. *Proc Lond Math Soc* 104:431–454
- Danilenko AI, Silva CE (2004) Multiple and polynomial recurrence for Abelian actions in infinite measure. *J Lond Math Soc* 69(2):183–200
- Danilenko AI, Solomko AV (2009) Infinite measure preserving flows with infinite ergodic index. *Colloq Math* 115:13–19
- Danilenko AI, Kosloff Z, Roy E (2022a) Nonsingular Poisson suspensions. *J d’Anal Math*, 146:741–790
- Danilenko AI, Kosloff Z, Roy E (2022b) Generic nonsingular Poisson suspension is of type III₁. *Ergodic Theory Dyn Syst* 42(4):1415–1445
- Day S, Grivna B, McCartney E, Silva CE (1999) Power weakly mixing infinite transformations. *New York J Math* 5:17–24
- del Junco A (1978) A simple measure-preserving transformation with trivial centralizer. *Pac J Math* 79:357–362
- del Junco A, Rudolph DJ (1987) On ergodic actions whose self-joinings are graphs. *Ergodic Theory Dyn Syst* 7: 531–557
- del Junco A, Şahin A (2009) Dyes’s theorem in the almost continuous category. *Israel J Math* 173:235–251
- del Junco A, Silva CE (1995) Prime type III_λ automorphisms: an instance of coding techniques applied to nonsingular maps. In: Takahashi Y (ed) *Fractals and dynamics* (Okayama/Kyoto, 1992). Plenum, New York, pp 101–115
- del Junco A, Silva CE (2003) On factors of nonsingular Cartesian products. *Ergodic Theory Dyn Syst* 23(5): 1445–1465
- Derriennic Y, Frączek K, Lemańczyk M, Parreau F (2008) Ergodic automorphisms whose weak closure of off-diagonal measures consists of ergodic self-joinings. *Colloq Math* 110:81–115
- Dixmier J (1969) *Les C^* -algèbres et leurs représentations*. Gauthier–Villars Editeur, Paris
- Dolgopyat D, Nándori P (2019) Infinite measure renewal theorem and related results. *Bull Lond Math Soc* 51: 145–167
- Dooley AN, Hagihara R (2012) Computing the critical dimensions of Bratteli-Vershik systems with multiple edges. *Ergodic Theory Dyn Syst* 32:103–117
- Dooley AH, Hamachi T (2003a) Nonsingular dynamical systems, Bratteli diagrams and Markov odometers. *Israel J Math* 138:93–123
- Dooley AH, Hamachi T (2003b) Markov odometer actions not of product type. *Ergodic Theory Dyn Syst* 23: 813–829
- Dooley AH, Mortiss G (2006) On the critical dimension and AC entropy for Markov odometers. *Monatsh Math* 149:193–213

- Dooley AH, Mortiss G (2007) The critical dimensions of Hamachi shifts. *Tohoku Math J.* 59(2):57–66
- Dooley AH, Mortiss G (2009) On the critical dimension of product odometers. *Ergodic Theory Dyn Syst* 29: 475–485
- Dooley AH, Klemes I, Quas AN (1998) Product and Markov measures of type III. *J Aust Math Soc Ser A* 65(1): 84–110
- Dye H (1959) On groups of measure-preserving transformations I. *Am J Math* 81:119–159, and II, *Am J Math* 85 (1963), 551–576
- Effros EG (1965) Transformation groups and C^* -algebras. *Ann Math* 81(2):38–55
- Eigen SJ (1981) On the simplicity of the full group of ergodic transformations. *Israel J Math* 40(3–4):345–349
- Eigen SJ (1982) The group of measure preserving transformations of $[0,1]$ has no outer automorphisms. *Math Ann* 259:259–270
- Eigen S, Hajian A, Halverson K (1998a) Multiple recurrence and infinite measure preserving odometers. *Israel J Math* 108:37–44
- Eigen S, Hajian A, Weiss B (1998b) Borel automorphisms with no finite invariant measure. *Proc Am Math Soc* 126:3619–3623
- Elyze M, Kastner A, Ortiz Rhoton J, Semenov V, Silva CE (2018) On conservative sequences and their application to ergodic multiplier problems. *Colloq Math* 151: 123–145
- Fedorov A (1985) Krieger’s theorem for cocycles, preprint
- Feldman J, Moore CC (1977) Ergodic equivalence relations, cohomology, and von Neumann algebras. I. *Trans Am Math Soc* 234:289–324
- Ferenczi S (1985) Systèmes de rang un gauche. *Ann Inst H Poincaré Probab Statist* 21:177–186
- Friedman NA (1970) Introduction to Ergodic Theory. Van Nostrand
- Friedman NA (1978) Mixing transformations in an infinite measure space, *Studies in probability and ergodic theory*. *Adv Math Suppl Stud* 2:167–184
- Furman A (2002) Random walks on groups and random transformations. In: *Handbook of dynamical systems*, vol 1A. North-Holland, Amsterdam, pp 931–1014
- Furstenberg H (1967) Disjointness in ergodic theory, minimal sets and diophantine approximation. *Math Syst Theory* 1:1–49
- Furstenberg H (1981) Recurrence in ergodic theory and combinatorial number theory. Princeton University Press, Princeton
- Furstenberg H, Glasner E (2010) Stationary dynamical systems. In: *Dynamical numbers interplay between dynamical systems and number theory*, *Contemporary mathematics*, vol 532. American Mathematical Society, pp 1–28
- Furstenberg H, Weiss B (1978) The finite multipliers of infinite ergodic transformations. In: *The structure of attractors in dynamical systems*, *Lecture notes in mathematics* 668. Springer, Berlin, pp 127–132
- Gaebler J, Kastner A, Silva CE, Xu X, Zhou Z (2018) Partially bounded transformations have trivial centralizers. *Proc Am Math Soc* 146:5113–5127
- Gårding L, Wightman AS (1954) Representation of anti-commutation relations. *Proc Natl Acad Sci U S A* 40: 617–621
- Giordano T, Skandalis G (1985a) Krieger factors isomorphic to their tensor square and pure point spectrum flows. *J Funct Anal* 64(2):209–226
- Giordano T, Skandalis G (1985b) On infinite tensor products of factors of type I_2 . *Ergodic Theory Dyn Syst* 5: 565–586
- Glasner E (1994) On the multipliers of $\mathcal{W}^\perp$. *Ergodic Theory Dyn Syst* 14:129–140
- Glasner E, Weiss B (2016) Weak mixing properties for non-singular actions. *Ergodic Theory Dyn Syst* 36: 2203–2217
- Glimm J (1961) Locally compact transformation groups. *Trans Am Math Soc* 101:124–138
- Golodets VY (1969) A description of the representations of anticommutation relations. *Uspehi Matemat Nauk* 24(4):43–64
- Golodets VY, Sinel’schikov SD (1983) Existence and uniqueness of cocycles of ergodic automorphism with dense range in amenable groups. Preprint FTINT AN USSR, pp 19–83
- Golodets VY, Sinel’schikov SD (1985) Locally compact groups appearing as ranges of cocycles of ergodic Z -actions. *Ergodic Theory Dyn Syst* 5:47–57
- Golodets VY, Sinel’schikov SD (1990) Amenable ergodic actions of groups and images of cocycles. (Russian). *Dokl Akad Nauk SSSR* 312(6):1296–1299
- Golodets VY, Sinel’schikov SD (1994) Classification and structure of cocycles of amenable ergodic equivalence relations. *J Funct Anal* 121:455–485
- Gouëzel S (2011) Correlation asymptotics from large deviations in dynamical systems with infinite measure. *Colloq Math* 125:193–212
- Grosser S, Moskowitz M (1971) Compactness conditions in topological groups. *J Reine Angew Math* 246:1–40
- Gruher K, Hines F, Patel D, Silva CE, Waelder R (2003) Power weak mixing does not imply multiple recurrence in infinite measure and other counterexamples. *New York J Math* 9:1–22
- Haddock B, Leng J, Silva CE (2022) Nonsingular transformations that are ergodic with isometric coefficients and not weakly doubly ergodic. *Indagationes Mathematicae* 33: 1297–1311
- Hajian AB, Kakutani S (1964) Weakly wandering sets and invariant measures. *Trans Am Math Soc* 110:136–151
- Halmos PR (1946) An ergodic theorem. *Proc Natl Acad Sci U S A* 32:156–161
- Halmos PR (1956) Lectures on ergodic theory. Publication of the Mathematical Society of Japan 3, Tokyo. Reprinted Chelsea Publishing Co., New York, 1960
- Hamachi T (1981a) The normalizer group of an ergodic automorphism of type III and the commutant of an ergodic flow. *J Funct Anal* 40:387–403

- Hamachi T (1981b) On a Bernoulli shift with nonidentical factor measures. *Ergodic Theory Dyn Syst* 1:273–283
- Hamachi T (1992) A measure theoretical proof of the Connes-Woods theorem on AT-flows. *Pac J Math* 154: 67–85
- Hamachi T, Kosaki H (1993) Orbital factor map. *Ergodic Theory Dyn Syst* 13:515–532
- Hamachi T, Osikawa M (1981) Ergodic groups of automorphisms and Krieger's theorems. *Seminar on Mathematical Science*, Keio University, 3
- Hamachi T, Osikawa M (1986) Computation of the associated flows of ITPFI₂ factors of type III₀. *Geometric methods in operator algebras* (Kyoto, 1983), Pitman research notes in mathematics series 123. Longman Scientific and Technical, Harlow, pp 196–210
- Hamachi T, Silva CE (2000) On nonsingular Chacon transformations. III *J Math* 44:868–883
- Hawkins JM (1982) Non-ITPFI diffeomorphisms. *Israel J Math* 42:117–131
- Hawkins JM (1983) Smooth type III diffeomorphisms of manifolds. *Trans Am Math Soc* 276:625–643
- Hawkins JM (1990a) Diffeomorphisms of manifolds with nonsingular Poincare flows. *J Math Anal Appl* 145(2): 419–430
- Hawkins JM (1990b) Properties of ergodic flows associated to product odometers. *Pac J Math* 141:287–294
- Hawkins J, Schmidt K (1982) On C^2 -diffeomorphisms of the circle which are of type III₁. *Invent Math* 66(3): 511–518
- Hawkins J, Silva CE (1991) Noninvertible transformations admitting no absolutely continuous σ -finite invariant measure. *Proc Am Math Soc* 111(2):455–463
- Hawkins J, Silva CE (1997/98) Characterizing mildly mixing actions by orbit equivalence of products. *New York J Math*, 3A. Proceedings of the New York Journal of Mathematics Conference, June 9–13, (1997), 99–115
- Hawkins J, Woods EJ (1984) Approximately transitive diffeomorphisms of the circle. *Proc Am Math Soc* 90(2):258–262
- Herman M (1979a) Construction de difféomorphismes ergodiques, preprint
- Herman M-R (1979b) Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations. (French). *Institute Hautes Etudes Science Publication Mathematics*, No. 49, 5–233
- Herman RH, Putnam IF, Skau CF (1992) Ordered Bratteli diagrams, dimension groups and topological dynamics. *Int J Math* 3(6):827–864
- Hochman M (2013) On the ratio ergodic theorem for group actions. *J Lond Math Soc* 88:465–482
- Hochman M (2019) Every Borel automorphism without finite invariant measures admits a two-set generator. *J Eur Math Soc (JEMS)* 21:271–317
- Hopf E (1937) *Ergodentheorie, Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol 3. Springer, Berlin, p 5
- Host B, Méla J-F, Parreau J-F (1986) Analyse harmonique des mesures. *Astérisque*, No. 135–136
- Host B, Méla J-F, Parreau F (1991) Nonsingular transformations and spectral analysis of measures. *Bull Soc Math France* 119:33–90
- Hurewicz W (1944) Ergodic theorem without invariant measure. *Ann Math* 45:192–206
- Inoue K (2004) Isometric extensions and multiple recurrence of infinite measure preserving systems. *Israel J Math* 140:245–252
- Ionescu Tulcea A (1965) On the category of certain classes of transformations in ergodic theory. *Trans Am Math Soc* 114:261–279
- James J, Koberda T, Lindsey K, Silva CE, Speh P (2008) Measurable sensitivity. *Proc Am Math Soc* 136: 3549–3559
- Janvresse E, de la Rue T (2012) Zero Krengel entropy does not kill Poisson entropy. *Annales de l'I.H.P. Probabilités et statistiques* 48:368–376
- Janvresse E, Meyerovitch T, de la Rue T, Roy E (2010) Poisson suspensions and entropy of infinite transformations. *Trans Am Math Soc* 362:3069–3094
- Janvresse E, de la Rue T, Roy E (2017) Poisson suspensions and SuShis. *Annales Scientifiques de l'École Normale Supérieure* 50:1301–1334
- Janvresse E, de la Rue T, Roy E (2018) Invariant measures for Cartesian powers of Chacon infinite transformation. *Israel J Math* 224:1–37
- Janvresse E, de la Rue T, Roy E (2019) Nearly finite Chacon transformation. *Ann H Lebesgue* 2:369–414
- Jaworski W (1994) Strongly approximatively transitive actions, the Choquet-Deny theorem, and polynomial growth. *Pac J Math* 165:115–129
- Johnson ASA, Şahin AA (2015) Directional recurrence for infinite measure preserving Z^d -actions. *Ergodic Theory Dyn Syst* 35:2138–2150
- Joita M, Munteanu RB (2014) A property of ergodic flows. *Stud Math* 225:249–258
- Kabanov JM, Lipcer RS, Sirjaev AN (1977) On the question of absolute continuity and singularity of probability measures. *Math USSR Sbornik* 33:203–221
- Kaimanovich VA, Vershik AM (1983) Random walks on groups: boundary and entropy. *Ann Probab* 11: 457–490
- Kakutani S (1948) On equivalence of infinite product measures. *Ann Math* 49:214–224
- Kakutani S, Parry W (1963) Infinite measure preserving transformations with “mixing”. *Bull Am Math Soc* 69: 752–756
- Katznelson Y (1977) Sigma-finite invariant measures for smooth mappings of the circle. *J Analyse Math* 31: 1–18
- Katznelson Y (1979) The action of diffeomorphism of the circle on the Lebesgue measure. *J Analyse Math* 36: 156–166
- Katznelson Y, Weiss B (1972) The construction of quasi-invariant measures. *Israel J Math* 12:1–4
- Katznelson Y, Weiss B (1991) The classification of nonsingular actions, revisited. *Ergodic Theory Dyn Syst* 11:333–348

- Keane M, Smorodinsky M (1979) Bernoulli schemes of the same entropy are finitarily isomorphic. *Ann Math* 109: 397–406
- King JL (1986) The commutant is the weak closure of the powers, for rank-1 transformations. *Ergodic Theory Dyn Syst* 6:363–384
- Kirillov AA (1978) Elements of the theory of representations. Nauka, Moscow
- Kosloff Z (2011) On a type III_λ Bernoulli shift. *Ergodic Theory Dyn Syst* 31:1727–1743
- Kosloff Z (2013) The zero-type property and mixing of Bernoulli shifts. *Ergodic Theory Dyn Syst* 33:549–559
- Kosloff Z (2014) On the K property for Maharam extensions of Bernoulli shifts and a question of Krengel. *Israel J Math* 199:485–506
- Kosloff Z (2017) A universal divergence rate for symmetric Birkhoff sums in infinite ergodic theory. *Trans Am Math Soc* 369:6373–6388
- Kosloff Z (2018) On manifolds admitting stable type III_λ Anosov diffeomorphisms. *J Modern Dyn* 13:251–270
- Kosloff Z (2019) Proving ergodicity via divergence of ergodic sums. *Stud Math* 248:191–215
- Kosloff Z (2021) Conservative Anosov diffeomorphisms of the two torus without an absolutely continuous invariant measure. *Ann Sci Ec Norm Supér* 54(4):69–131
- Kosloff Z, Soo T (2021) The orbital equivalence of Bernoulli actions and their Sinai factors. *J Mod Dyn* 17:145–182
- Kosloff Z, Soo T (2022) Some factors of nonsingular Bernoulli shifts. *Stud Math* 262(1):23–43
- Krengel U (1967) Entropy of conservative transformations. *Z Wahrscheinlichkeitstheorie und Verw Gebiete* 7:161–181
- Krengel U (1969) Darstellungssätze für Strömungen und Halbströmungen. II. *Math Ann* 182:1–39
- Krengel U (1970) Transformations without finite invariant measure have finite strong generators. In: *Contributions to Ergodic theory and probability (Proceedings of the Conference on Ohio State University, Columbus, 1970)*. Springer, Berlin, pp 133–157
- Krengel U (1976) On Rudolph's representation of aperiodic flows. *Ann Inst H Poincaré Sect B (NS)* 12(4):319–338
- Krengel U (1985) *Ergodic theorems*. de Gruyter Studies in Mathematics, Berlin
- Krengel U, Sucheston L (1969) On mixing in infinite measure spaces. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 13:150–164
- Krickeberg K (1967) Strong mixing properties of Markov chains with infinite invariant measure. In: *Proceedings of the Fifth Berkeley symposium on mathematical statistics and probability (Berkeley, 1965/66)*, Vol. II: *Contributions to probability theory, Part 2*. University of California Press, Berkeley, pp 431–446
- Krieger W (1969) On nonsingular transformations of a measure space, I, II. *Z Wahrscheinlichkeitstheorie und Verw Gebiete* 11:83–119
- Krieger W (1970) On the Araki-Woods asymptotic ratio set and nonsingular transformations of a measure space. In: *Contributions to Ergodic theory and probability (Proceedings of the conference on Ohio State University, Columbus, 1970)*. Lecture notes in mathematics, vol 160. Springer, Berlin, pp 158–177
- Krieger W (1972) On the infinite product construction of nonsingular transformations of a measure space. *Invent Math* 15:144–163; and erratum in 26 (1974), 323–328
- Krieger W (1976a) On Borel automorphisms and their quasi-invariant measures. *Math Z* 151:19–24
- Krieger W (1976b) On ergodic flows and isomorphism of factors. *Math Ann* 223:19–70
- Kubo I (1969) Quasi-flows. *Nagoya Math J* 35:1–30
- Landesberg O, Lindenstrauss E (2022) On Radon measures invariant under horospherical flows on geometrically infinite quotients. *Int Math Res Not IMRN*:11602–11641
- Ledrappier F, Sarig O (2007) Invariant measures for the horocycle flow on periodic hyperbolic surfaces. *Israel J Math* 160:281–315
- Lehrer E, Weiss B (1982) An ϵ -free Rokhlin lemma. *Ergodic Theory Dyn Syst* 2:45–48
- Lemańczyk M, Parreau F (2003) Rokhlin extensions and lifting disjointness. *Ergodic Theory Dyn Syst* 23: 1525–1550
- Lemańczyk M, Parreau F, Thouvenot J-P (2000) Gaussian automorphisms whose ergodic self-joinings are Gaussian. *Fundam Math* 164:253–293
- Lemańczyk M, Lesigne E, Skrenty D (2001) Multiplicative Gaussian cocycles. *Aequationes Math* 61:162–178
- Lenci M (2010) On infinite-volume mixing. *Commun Math Phys* 298:485–514
- Lenci M (2013) Exactness, K -property and infinite mixing. *Publ Mat Urug* 14:159–170
- Lenci M (2017) Uniformly expanding Markov maps of the real line: exactness and infinite mixing. *Discrete Contin Dyn Syst* 37:3867–3903
- Loh I, Silva CE (2017) Strict doubly ergodic infinite transformations. *Dyn Syst* 32(4):519–543
- Loh I, Silva CE, Athiwaratkun B (2018) Infinite symmetric ergodic index and related examples in infinite measure. *Stud Math* 243:101–115
- Mackey GW (1966) Ergodic theory and virtual group. *Math Ann* 166:187–207
- Maharam D (1964) Incompressible transformations. *Fund Math* LVI:35–50
- Mandrekar V, Nadkarni M (1969) On ergodic quasi-invariant measures on the circle group. *J Funct Anal* 3:157–163
- Mansfield DF, Dooley AN (2017) The critical dimension for G -measures. *Ergodic Theory Dyn Syst* 37:824–836
- Marrakchi A, Vaes S (2022) Nonsingular Gaussian actions: beyond the mixing case. *Adv Math* 397:108190, 62 p
- Matui H (2002) Topological orbit equivalence of locally compact Cantor minimal systems. *Ergodic Theory Dyn Syst* 22:1871–1903
- Méla J-F (1983) Groupes de valeurs propres des systèmes dynamiques et sous-groupes saturés du cercle. *C R Acad Sci Paris Sér I Math* 296(10):419–422
- Melbourne I (2015) Mixing for invertible dynamical systems with infinite measure. *Stoch Dyn* 15(2): 1550012, 25

- Melbourne I, Terhesiu D (2012) Operator renewal theory and mixing rates for dynamical systems with infinite measure. *Invent Math* 189:61–110
- Meyerovitch T (2007) On multiple and polynomial recurrent extensions of infinite measure preserving transformations, preprint. ArXiv: <http://arxiv.org/abs/math/0703914>
- Meyerovitch T (2013) Ergodicity of Poisson products and applications. *Ann Probab* 41:3181–3200
- Milnor J (1988) On the entropy geometry of cellular automata. *Complex Syst* 2:357–385
- Moore CC (1967) Invariant measures on product spaces. In: *Proceedings of the Fifth Berkeley symposium*, pp 447–459
- Moore CC (1982) Ergodic theory and von Neumann algebras. In: *Proceedings of the symposium on pure mathematics*, 38. American Mathematical Society, Providence, pp 179–226
- Moore CC, Schmidt K (1980) Coboundaries and homomorphisms for nonsingular actions and a problem of H. Helson. *Proc Lond Math Soc* 40(3):443–475
- Mortiss G (2000) A non-singular inverse Vitali lemma with applications. *Ergodic Theory Dyn Syst* 20:1215–1229
- Mortiss G (2002) Average co-ordinate entropy. *J Aust Math Soc* 73:171–186
- Mortiss G (2003) An invariant for nonsingular isomorphism. *Ergodic Theory Dyn Syst* 23:885–893
- Munteanu RB (2012) A non-product type non-singular transformation which satisfies Krieger's Property A. *Israel J Math* 190:307–324
- Nadkarni MG (1979) On spectra of nonsingular transformations and flows. *Sankhya Ser A* 41(1–2):59–66
- Nadkarni MG (1998) *Spectral theory of dynamical systems*, Birkhäuser Advanced Texts: Basler Lehrbücher. Birkhäuser Verlag, Basel
- Neretin YA (1996) *Categories of symmetries and infinite-dimensional groups*. Oxford University Press, New York
- Ornstein D (1960) On invariant measures. *Bull Am Math Soc* 66:297–300
- Ornstein D (1972) On the root problem in ergodic theory. In: *Proceedings of the Sixth Berkeley symposium on mathematical statistics and probability*. University of California Press, pp 347–356
- Osikawa M (1977/78) Point spectra of nonsingular flows. *Publ Res Inst Math Sci* 13:167–172
- Osikawa M (1988) Ergodic properties of product type odometers. *Springer Lect Notes Math* 1299:404–414
- Osikawa M, Hamachi T (1971) On zero type and positive type transformations with infinite invariant measures. *Mem Facul Sci, Kyushu Univ* 25:280–295
- Parreau F, Roy E (2008) Poisson joinings of Poisson suspensions, preprint
- Parry W (1963) An ergodic theorem of information theory without invariant measure. *Proc Lond Math Soc* 13(3): 605–612
- Parry W (1965) Ergodic and spectral analysis of certain infinite measure preserving transformations. *Proc Am Math Soc* 16:960–966
- Parry W (1966) Generators and strong generators in ergodic theory. *Bull Am Math Soc* 72:294–296
- Parry W (1969) *Entropy and generators in ergodic theory*. W. A. Benjamin, Inc., New York/Amsterdam
- Parthasarathy KR, Schmidt K (1977) On the cohomology of a hyperfinite action. *Monatsh Math* 84(1):37–48
- Ramsay A (1971) Virtual groups and group actions. *Adv Math* 6:243–322
- Rokhlin VA (1965) Generators in ergodic theory. II. (Russian. English summary). *Vestnik Leningrad Univ* 20(13):68–72
- Rosinsky J (1995) On the structure of stationary stable processes. *Ann Probab* 23:1163–1187
- Roy E (2005) Mesures de Poisson, infinie divisibilité et propriétés ergodiques, Thèse de doctorat de l'Université Paris 6
- Roy E (2007) Ergodic properties of Poissonian ID processes. *Ann Probab* 35:551–576
- Roy E (2009) Poisson suspensions and infinite ergodic theory. *Ergodic Theory Dyn Syst* 29:667–683
- Roy E (2010) Poisson-Pinsker factor and infinite measure preserving group actions. *Proc Am Math Soc* 138: 2087–2094
- Rudolph DJ (1985) Restricted orbit equivalence. *Mem Am Math Soc* 323
- Rudolph D, Silva CE (1989) Minimal self-joinings for nonsingular transformations. *Ergodic Theory Dyn Syst* 9:759–800
- Ryzhikov VV (1993) Factorization of an automorphism of a full Boolean algebra into the product of three involutions. (Russian). *Mat Zametki* 54(2):79–84, 159; translation in *Math. Notes* 54 (1993), no. 1–2, 821–824 (1994)
- Ryzhikov VV (2014) On the asymmetry of multiple asymptotic properties of ergodic actions. *Math Notes* 96:416–422
- Ryzhikov VV, Thouvenot J-P (2015) On the centralizer of an infinite mixing rank-one transformation. *Funct Anal Appl* 49:230–233
- Sachdeva U (1971) On category of mixing in infinite measure spaces. *Math Syst Theory* 5:319–330
- Samorodnitsky G (2005) Null flows, positive flows and the structure of stationary symmetric stable processes. *Ann Probab* 33:1782–1803
- Sarig O (2004) Invariant measures for the horocycle flows on Abelian covers. *Invent Math* 157:519–551
- Schmidt K (1977a) Cocycles on ergodic transformation groups. *Macmillan lectures in mathematics*, vol 1. Macmillan Company of India, Ltd., Delhi
- Schmidt K (1977b) Infinite invariant measures in the circle. *Symp Math* 21:37–43
- Schmidt K (1982) Spectra of ergodic group actions. *Israel J Math* 41(1–2):151–153
- Schmidt K (1984) On recurrence. *Z Wahrscheinlichkeitstheorie verw Gebiete* 68:75–95
- Schmidt K, Walters P (1982) Mildly mixing actions of locally compact groups. *Proc Lond Math Soc* 45: 506–518

- Shelah S, Weiss B (1982) Measurable recurrence and quasi-invariant measures. *Israel J Math* 43:154–160
- Silva CE (1988) On μ -recurrent nonsingular endomorphisms. *Israel J Math* 61:1–13
- Silva CE, Thieullen P (1991) The subadditive ergodic theorem and recurrence properties of Markovian transformations. *J Math Anal Appl* 154(1):83–99
- Silva CE, Thieullen P (1995) A skew product entropy for nonsingular transformations. *J Lond Math Soc* 52(2):497–516
- Silva CE, Witte D (1992) On quotients of nonsingular actions whose self-joinings are graphs. *Int J Math* 5:219–237
- Thaler M (1998) The Dynkin-Lamperti arc-sine laws for measure preserving transformations. *Trans Am Math Soc* 350(11):4593–4607
- Thaler M, Zweimüller R (2006) Distributional limit theorems in infinite ergodic theory. *Probab Theory Relat Fields* 135(1):15–52
- Thouvenot J-P (1995) Some properties and applications of joinings in ergodic theory. In: *Ergodic theory and its connections with harmonic analysis* (Alexandria, 1993), London Mathematical Society, Lecture Note Series, 205. Cambridge University Press, Cambridge, pp 207–235
- Ullman D (1987) A generalization of a theorem of Atkinson to non-invariant measures. *Pac J Math* 130:187–193
- Vaes S, Wahl J (2018) Bernoulli actions of type III_λ and L^2 -cohomology. *Geom Funct Anal* 28:518–562
- Vershik AM (1983) Many valued mappings with invariant measure (polymorphisms) and Markov processes. *J Sov Math* 23:2243–2266
- Vershik AM, Kerov SV (1985) Locally semisimple algebras. Combinatorial theory and K_0 -functor. *Modern Prob Math* 26:3–56
- Yuasa H (2013) Uniform sets for infinite measure-preserving systems. *J d'Anal Math* 120:333–356
- Yuasa H (2020) A relative, strictly ergodic model theorem for infinite measure-preserving systems. *J d'Anal Math* 140:591–616
- Zimmer RJ (1977) Random walks on compact groups and the existence of cocycles. *Israel J Math* 26:84–90
- Zimmer RJ (1978) Amenable ergodic group actions and an application to Poisson boundaries of random walks. *J Funct Anal* 27:350–372

Sarnak's Conjecture from the Ergodic Theory Point of View

Joanna Kułaga-Przymus and Mariusz Lemańczyk
Faculty of Mathematics and Computer Science,
Nicolaus Copernicus University, Toruń, Poland

Article Outline

Glossary
Definition of the Subject
Introduction
Chowla Conjecture
Sarnak's Conjecture
Arithmetic Properties of the Möbius Function
Future Directions
Bibliography

Glossary

Aperiodicity We say that $u : \mathbb{N} \rightarrow \mathbb{C}$ is *aperiodic* whenever u has a mean, equal to zero, along each arithmetic progression: $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n \leq N} u(an + b) = 0$. Many classical multiplicative functions are aperiodic, including μ and λ .

For a parameter $N \geq 1$, a distance between $u, v : \mathbb{N} \rightarrow \mathbb{U}$ is defined as
$$\mathbb{D}(u, v; N) := \left(\sum_{p \in \mathbb{P}, p \leq N} \frac{1 - \operatorname{Re}(u(p)\overline{v(p)})}{p} \right)^{1/2}.$$

We say that $u : \mathbb{N} \rightarrow \mathbb{U}$ is *strongly aperiodic* (Matomäki et al. 2015), whenever $M(u \cdot \chi; N) := \min_{|t| \leq N} \mathbb{D}(u \cdot \chi, n^{it}; N)^2 \rightarrow \infty$ as $N \rightarrow \infty$ for every Dirichlet character χ (i.e., for every periodic, completely multiplicative function). Strong aperiodicity implies aperiodicity. The converse is not in general true (see Theorem B.1 in (Matomäki et al. 2015)), but it is true for (bounded) real valued multiplicative functions

(see Appendix C in (Matomäki et al. 2015)). In particular, μ and λ are strongly aperiodic.

Arithmetic function A sequence of complex numbers is usually denoted by $u = (u_n)$. But if such a sequence is, in some sense, important from number theory point of view, one speaks about an *arithmetic function* and rather writes $u : \mathbb{N} \rightarrow \mathbb{C}$, $u = (u(n))$.

An arithmetic function u is said to be *multiplicative*, whenever $u(1) = 1$ and $u(m \cdot n) = u(m) \cdot u(n)$ for any choice of coprime $m, n \in \mathbb{N}$. The prominent examples of multiplicative functions are the Möbius function μ and the Liouville function λ . The Möbius function $\mu : \mathbb{N} \rightarrow \{-1, 0, 1\}$ is defined by $\mu(p_1 \dots p_k) = (-1)^k$ for different prime numbers $p_1, \dots, p_k$ (in what follows, the set of primes is denoted by $\mathbb{P}$), $\mu(1) = 1$ and $\mu(n) = 0$ for all non-square-free numbers. The Liouville function $\lambda : \mathbb{N} \rightarrow \{-1, 1\}$ is given by $\lambda(n) = (-1)^{i_1 + \dots + i_k}$ for $n = p_1^{i_1} \dots p_k^{i_k}$ with $p_1, \dots, p_k \in \mathbb{P}$ and $i_1, \dots, i_k \in \mathbb{N}$. Clearly $\mu = \lambda \cdot \mu^2$, where μ^2 is nothing but the characteristic function of the set S of square-free numbers. In fact, λ is completely multiplicative, that is, $\lambda(m \cdot n) = \lambda(m) \cdot \lambda(n)$ for any choice of $m, n \in \mathbb{N}$. We extend both, μ and λ , to negative coordinates symmetrically.

Completely deterministic point We say that point $x \in X$ is *completely deterministic* (Weiss 1971) (see also (Kamae 1973)) if for any $v \in V(x)$, we have $h(T, X, \mathcal{B}(X), v) = 0$. By the variational principle, $h_{\text{top}}(T, X) = 0$ if and only if all points of X are completely deterministic.

Entropy There are two basic notions of *entropy*: topological and measure-theoretic. We skip the definitions and refer the reader, for example, to (Downarowicz 2011). For a topological dynamical system (T, X) its topological entropy will be denoted by $h_{\text{top}}(T, X)$ and for a measure-theoretic dynamical system $(T, X, \mathcal{B}, v)$ the corresponding measure-theoretic entropy will

be denoted by $h(T, X, \mathcal{B}, v)$. The basic connection between them is the variational principle:

$$h_{\text{top}}(T, X) = \sup_{v \in M(T, X)} h(T, X, \mathcal{B}(X), v) = \sup_{v \in M^e(T, X)} h(T, X, \mathcal{B}(X), v).$$

Furstenberg system Let $u \in \mathbb{A}^{\mathbb{Z}}$. For each $v \in V(u)$, the system $(S, X_u, \mathcal{B}(X_u), v)$ is called a *Furstenberg system* of u . For each $v \in V^{\log}(u)$, the system $(S, X_u, \mathcal{B}(X_u), v)$ is called a *logarithmic Furstenberg system* of u .

For each $|u| \in \mathbb{A}^{\mathbb{Z}}$, one can consider $|u| \in [0, 1]^{\mathbb{Z}}$. Then $(S, X_{|u|})$ is a topological factor of (S, X_u) (the map $\pi : X_u \rightarrow X_{|u|}$ given by $\pi(x) = |x|$, understood coordinatewise, is equivariant with S). For $v \in V(u)$, we have $\pi_*(v) \in V(|u|)$, where $\pi_*(v)$ stands for the image of v via π . Moreover, the Furstenberg system $(S, X_u, \mathcal{B}(X_u), v)$ is an extension of $(S, X_{|u|}, \mathcal{B}(X_{|u|}), \pi_*(v))$. In particular, if $V(|u|)$ is a singleton ($|u|$ is a generic point), then all Furstenberg systems of u have $(S, X_{|u|}, \mathcal{B}(X_{|u|}), v)$ (where v is the unique member of $V(|u|)$) as their factor.

Generic point Assume that (T, X) is a topological dynamical system. We say that $x \in X$ is a *generic point* for a Borel measure v on X , whenever the ergodic theorem holds for T at x for any continuous function $f \in C(X)$, that is, $\frac{1}{N} \sum_{n \leq N} f(T^n x) \rightarrow \int f dv$. In other words, the empirical measures $\frac{1}{N} \sum_{n \leq N} \delta_{T^n x}$ converge to v in the weak- $*$ -topology. If this convergence holds along a subsequence, we say that x is *quasi-generic* for v . We denote by $V(x)$ the set of all $v \in M(T, X)$ for which x is quasi-generic ($\emptyset \neq V(x) \subset M(T, X)$ by compactness). We say that $x \in X$ is *logarithmically generic* for v , whenever $\frac{1}{L_N} \sum_{n \leq L_N} \frac{1}{n} \delta_{T^n x}$, with $L_N = \sum_{n \leq N} 1/n$, converges to v . No harm arises if in what follows we replace L_N by $\log N$. Finally, we say that $x \in X$ is *logarithmically quasi-generic* for v whenever this convergence holds along a subsequence and we denote the set of all measures for which x is logarithmically quasi-generic by $V^{\log}(x)$ ($\emptyset \neq V^{\log}(x) \subset M(T, X)$).

Invariant measure Given a topological dynamical system (T, X) , the set of probability Borel

T -invariant measures is denoted by $M(T, X)$. The subset of ergodic measures (which is always non-empty) is denoted by $M^e(T, X)$. Each $v \in M(T, X)$ gives rise to a measure-theoretic dynamical system $(T, X, \mathcal{B}(X), v)$, where $\mathcal{B}(X)$ stands for the σ -algebra of Borel subsets of X . With the weak- $*$ -topology, $M(T, X)$ becomes a compact metrizable space. (T, X) is called *uniquely ergodic* if $|M(T, X)| = 1$.

Joinings of measure-theoretic dynamical systems Assume that $(R_i, Z_i, \mathcal{C}_i, \kappa_i)$ is a measure-theoretic dynamical system, $i = 1, 2$. Each $R_1 \times R_2$ -invariant measure ρ on $\mathcal{C}_1 \otimes \mathcal{C}_2$ projecting on κ_1 and κ_2 , respectively, is called a *joining* of the automorphisms R_1 and R_2 . The set of joinings between R_1 and R_2 is denoted by $J(R_1, R_2)$ and each $\rho \in J(R_1, R_2)$ yields a (new) measure-theoretic dynamical system $(R_1 \times R_2, Z_1 \times Z_2, \mathcal{C}_1 \otimes \mathcal{C}_2, \rho)$. When R_1 and R_2 are both ergodic, then the set $J(R_1, R_2)$ of ergodic joinings between R_1 and R_2 is non-empty. If $J(R_1, R_2) = \{\kappa_1 \otimes \kappa_2\}$, then R_1 and R_2 are called *disjoint* (in the sense of Furstenberg).

Let (T, X) be a topological dynamical system. Fix $x \in X$ and let u be an arithmetic function bounded by 1, that is, $u : \mathbb{N} \rightarrow \mathbb{U}$. Each accumulation point ρ of $\frac{1}{N} \sum_{n \leq N} \delta_{(T^n x, S^n u)}$ is a $(T \times S)$ -invariant measure on $X \times \mathbb{U}^{\mathbb{Z}}$. Obviously, it is a *joining* of its projections onto both coordinates. Hence, ρ is a joining of $(T, X, \mathcal{B}(X), \rho|_X)$ with a Furstenberg system of u .

Measure-theoretic dynamical system Let $(Z, \mathcal{C}, \kappa)$ be a standard probability Borel space. Assume $R : X \rightarrow X$ is invertible (a.e.), bi-measurable, and κ -preserving. Then R is called an *automorphism* or (invertible) *measure-preserving transformation* and the quadruple $(R, Z, \mathcal{C}, \kappa)$ is a *measure-theoretic dynamical system*.

Möbius orthogonality We say that a (topological) dynamical system (T, X) is *Möbius orthogonal* if

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n \leq N} f(T^n x) \mu(n) = 0 \quad (3)$$

for each $f \in C(X)$ and $x \in X$.

Nilrotation Let G be a connected, simply connected nilpotent Lie group and $\Gamma \subset G$ a lattice (a discrete, cocompact subgroup). For any $g_0 \in G$ we define $T_{g_0}(g\Gamma) := g_0g\Gamma$. Then the topological system $(T_{g_0}, G/\Gamma)$ is called a *nilrotation*.

Orthogonality of sequences Suppose that one of sequences $(u_n), (v_n) \subset \mathbb{C}$ is of zero mean. We say that $(u_n), (v_n)$ are *orthogonal*, whenever $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n \leq N} u_n \bar{v}_n = 0$. We say that sequences $(u_n), (v_n) \subset \mathbb{C}$ are *orthogonal on short intervals*, whenever $\lim_{K \rightarrow \infty} \frac{1}{b_K} \sum_{k \leq K} \left| \sum_{b_k \leq n < b_{k+1}} u_n \bar{v}_n \right| = 0$ for any sequence (b_k) of natural numbers with $b_{k+1} - b_k \rightarrow \infty$. Clearly, orthogonality on short intervals implies orthogonality. We can consider u_n being an element of a Banach space, while in the dynamical context one often takes $u_n = f(T^n x_k)$, whenever $n \in [b_k, b_{k+1})$ (with $x_k \in X$ and $f \in C(X)$). Moreover, often (v_n) will be in fact a multiplicative function (and then we rather write $(v(n))$).

Quasi-genericity versus logarithmic quasi-genericity Obviously, if $|V(x)| = 1$, then $V(x) = V^{\log}(x)$, but in general the sets $V(x)$ and $V^{\log}(x)$ can be even disjoint. However, as shown in (Gomilko et al. 2018),

$$V^{\log}(x) \cap M^e(T, X) \subset V(x). \quad (1)$$

Moreover, using an idea from Tao (n.d.), it has been proved in Gomilko et al. (2020) that

$$\begin{aligned} &\text{if } V^{\log}(x) = \{v\} \text{ and } v \text{ is ergodic, then} \\ &\lim_{k \rightarrow \infty} \frac{1}{N_k} \sum_{n \leq N_k} \delta_{T^n x} = v, \text{ for a subset} \\ &\quad \{N_k : k \geq 1\} \text{ whose} \\ &\quad \text{logarithmic density is 1.} \end{aligned} \quad (2)$$

Subshift For any closed subset $\mathbb{A} \subset \mathbb{U} := \{z \in \mathbb{C} : |z| \leq 1\}$ (most often $\mathbb{A}$ will be finite), let $S : \mathbb{A}^{\mathbb{Z}} \rightarrow \mathbb{A}^{\mathbb{Z}}$ be the left shift, that is, for each $x = (x_n)_{n \in \mathbb{Z}} \in \mathbb{A}^{\mathbb{Z}}$, $Sx = y$, where $y_n = x_{n+1}$ for $n \in \mathbb{Z}$. Each closed and S -invariant subset $X \subset \mathbb{A}^{\mathbb{Z}}$ is called a *subshift*

and the corresponding dynamical system is (S, X) . For each $u \in \mathbb{A}^{\mathbb{Z}}$, we define a subshift $X_u \subset \mathbb{A}^{\mathbb{Z}}$ as the orbit closure of u under S . Similarly, when $u \in \mathbb{A}^{\mathbb{N}}$ ($\mathbb{N} = \{1, 2, \dots\}$), we can extend u symmetrically, by setting $u(-n) := u(n)$ (and $u(0) = 0$) for each $n \geq 1$ and define again the corresponding subshift X_u . Finally, we will denote by F the continuous function on $X \subset \mathbb{A}^{\mathbb{Z}}$, defined by $F(x) = x_0$.

Topological dynamical system Let X be a compact metric space and let T be a homeomorphism of X . Then (T, X) is called a *topological dynamical system*.

Uniquely ergodic model of a measure-theoretic dynamical system Given a measure-theoretic (ergodic) dynamical system (R, Z, C, κ) and a uniquely ergodic topological system (T, X) with $M(T, X) = \{v\}$ (v is necessarily ergodic), one says that (T, X) is a *uniquely ergodic model* of the automorphism R if the measure-theoretic systems (R, Z, C, κ) and $(T, X, \mathcal{B}(X), v)$ are measure-theoretically isomorphic. Each ergodic automorphism R has a uniquely ergodic model.

Definition of the Subject

Sarnak in 2010 formulated a now celebrated conjecture on the Möbius orthogonality. It states that each topological dynamical system (T, X) of zero entropy is Möbius orthogonal, that is, whenever $h_{\text{top}}(T, X) = 0$,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n \leq N} f(T^n x) \mu(n) = 0 \quad (4)$$

for each $f \in C(X)$ and $x \in X$.

It has become a bridge between analytic number theory and dynamics. In particular, it keeps stimulating a quick development of disjointness theory in dynamics originated by Furstenberg in 1967, with potential applications, for example, in number theory. In this entry, we will focus on ergodic-theoretic aspects of recent progress in the area. Unless we need this for a historical reason, we rather avoid describing numerous particular classes

of topological systems in which Möbius disjointness has been shown. We concentrate on main ideas, and describe some general results.

Introduction

The entry is organized as follows. In sections “[Chowla Conjecture](#)” and “[Sarnak's Conjecture](#),” respectively, we discuss Chowla and Sarnak's conjectures from the ergodic-theoretic viewpoint, including dynamical interpretations of purely number-theoretic statements and the main strategies used to attack Sarnak's conjecture. Section “[Arithmetic Properties of the Möbius Function](#)” is a survey of results on Sarnak's conjecture, arranged with respect to the properties of the Möbius function that come into play in the proof. Finally, in section “[Future Directions](#),” we state some open problems.

Chowla Conjecture

(C)

The Chowla conjecture deals with higher order correlations of the Möbius function and asserts that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n \leq N} \mu^{j_0}(n+k_0) \mu^{j_1}(n+k_1) \dots \mu^{j_r}(n+k_r) = 0 \quad (\text{C})$$

whenever $0 \leq k_0 < \dots < k_r, j_s \in \{1, 2\}$ not all equal to 2, $r \geq 0$ (in Chowla (1965), the Chowla conjecture is formulated for the Liouville function. We follow Sarnak (n.d.). For a discussion on an equivalence of the Chowla conjecture with μ and λ , see (Ramaré 2018)). The above statement can be translated into dynamical language. To see this, consider first $|\mu| = \mu^2$, that is, the characteristic function of the set S of square-free numbers. As a member of $\{0, 1\}^{\mathbb{Z}}$, it is well-known to be a generic point for an ergodic measure, namely for the so-called Mirsky measure ν_S (also denoted as ν_{μ^2}), given by the ordinary average frequencies of blocks, see, for example, El Abdalaoui et al. (2017a). (C) is equivalent to the following:

the point $\mu \in \{-1, 0, 1\}^{\mathbb{Z}}$ is generic for the relatively independent extension $\widehat{\nu}_S$ of ν_S , via the natural map

$$\{-1, 0, 1\}^{\mathbb{Z}} \ni (x_n)_{n \geq 1} \mapsto (x_n^2)_{n \geq 1} \in \{0, 1\}^{\mathbb{Z}}. \quad (5)$$

The measure $\widehat{\nu}_S$ is given by the following condition: for each block C over the alphabet $\{-1, 0, 1\}$, we have $\widehat{\nu}_S(C) = 2^{-k} \nu_S(C^2)$, where C^2 is obtained from C by taking the square (or, equivalently, absolute value) at each coordinate and k is the cardinality of the support of C . To see this, we use the following:

Remark 4.1 The span of the family of continuous functions $\{F^{j_0} \circ S^{k_0} \cdot F^{j_1} \circ S^{k_1} \cdot \dots \cdot F^{j_\ell} \circ S^{k_\ell} : j_i \geq 0, k_i \in \mathbb{Z}\}$ forms an algebra that distinguishes points. It follows directly from the Stone-Weierstrass theorem that the values of integrals $\int F^{k_1} \circ S^{r_1} \cdot F^{k_2} \circ S^{r_2} \cdot \dots \cdot F^{k_\ell} \circ S^{r_\ell} d\kappa$ determine a measure κ .

Now, it suffices to look at measures $\kappa \in V(\mu)$ and compare the value of the integrals $\int F^{k_1} \circ S^{r_1} \cdot F^{k_2} \circ S^{r_2} \cdot \dots \cdot F^{k_\ell} \circ S^{r_\ell} d\kappa$ with the corresponding values of $\int F^{k_1} \circ S^{r_1} \cdot F^{k_2} \circ S^{r_2} \cdot \dots \cdot F^{k_\ell} \circ S^{r_\ell} d\widehat{\nu}_S$. We will also use the fact that

$$\widehat{\nu}_S \in M^e(S, X_\mu). \quad (6)$$

The simplest instance of Chowla conjecture, that is, $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n \leq N} \mu(n) = 0$, is known to hold and it is equivalent to the Prime Number Theorem, as shown by Landau. Moreover, if exactly one of the exponents j_i is odd, then (C) also holds (in Ram Murty and Vatwani (2018) a quantitative version of this fact has been proved). We also have the following conditional result of Frantzikinakis on (C):

Theorem 4.2 (Frantzikinakis 2017) *Assume that $V(\lambda) = \{\kappa\}$ and κ is ergodic. Then (C) holds for λ (i.e., κ equals the $\frac{1}{2}$ -Bernoulli measure on $\{-1, 1\}^{\mathbb{Z}}$. Cf. Theorem 4.4).*

(C) vs. (C_{log})

Analogously, one can study the logarithmic Chowla conjecture (Tao 2017), asserting that

$$\lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{n \leq N} \frac{1}{n} \mu^{j_0}(n+k_0) \mu^{j_1}(n+k_1) \dots$$

$$\mu^{j_r}(n+k_r) = 0, \quad (\text{C}_{\log})$$

with all parameters as before. Again, this can be translated to the dynamical language, using the notion of a logarithmically generic point. Thus, using (6), (1), and (2), we have the following:

Corollary 4.3 (cf. Gornilko et al. (2018); Tao (n.d.)). $(\text{C}_{\log})$ implies (C) along a subsequence of full logarithmic density.

More on $(\text{C}_{\log})$

We begin this section with a conditional result of Frantzikinakis on $(\text{C}_{\log})$ (cf. Theorem 4.2):

Theorem 4.4 (Frantzikinakis 2017) *Assume that $\kappa \in \mathcal{V}^{\log}(\lambda)$ is an ergodic measure. Then $(\text{C}_{\log})$ holds for λ along the same subsequence for which λ is quasi-generic for κ .*

The above result is already formulated in the language of (logarithmically) quasi-generic points. We pass now to number-theoretic results, with their ergodic-theoretic consequences.

Theorem 4.5 (Tao 2016)

$(\text{C}_{\log})$ holds for $r = 1$ (for λ in place of μ), that is, for each $0 \neq h \in \mathbb{Z}$, we have

$$\lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{n \leq N} \frac{\lambda(n) \lambda(n+h)}{n} = 0.$$

In fact, Tao in (Tao 2016) proves a stronger result, being an instance of logarithmically averaged Elliott conjecture. One of the consequences is that the analogue of Theorem 4.5 holds for μ :

Corollary 4.6 (Tao 2016)

$(\text{C}_{\log})$ holds for $r = 1$, that is, for each $0 \neq h \in \mathbb{Z}$, we have

$$\lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{n \leq N} \frac{\mu(n) \mu(n+h)}{n} = 0.$$

This, in turn, has the following interpretation in terms of spectral measures:

Corollary 4.7 (Ferenczi et al. 2018)

For each logarithmic Furstenberg system $(S, X_{\mu}, \mathcal{B}(X_{\mu}), \kappa)$ of μ , the spectral measure σ_F of F is Lebesgue. The same holds for $\kappa \in \mathcal{V}^{\log}(\lambda)$.

Moreover, we have the following:

Theorem 4.8 (Tao and Teräväinen 2018, 2019b)

$(\text{C}_{\log})$ holds for odd order correlations, that is, for all even values of r .

Again, we want to interpret this from the dynamical viewpoint:

Corollary 4.9 For each logarithmic Furstenberg system $(S, X_{\mu}, \mathcal{B}(X_{\mu}), \kappa)$, the element $Ry := -y$ preserves measure κ and commutes with the shift S , hence R belongs to the centralizer of S .

To see that the above statement is true, it suffices to take $\kappa \in \mathcal{V}^{\log}(\mu)$ and check that we also have $\kappa \in \mathcal{V}^{\log}(-\mu)$. For this, one uses the family of continuous functions $\{F^{j_0} \circ S^{k_0} \cdot F^{j_1} \circ S^{k_1} \cdot \dots \cdot F^{j_\ell} \circ S^{k_\ell} : j_i \geq 0, k_i \in \mathbb{Z}\}$. Then, for even correlations the desired equalities are obvious and for odd correlations one applies Theorem 4.8.

Finally, as a consequence of Corollary 4.3, we have the following:

Corollary 4.10 (Gornilko et al. 2018).

Suppose that $(\text{C}_{\log})$ holds. Then $\widehat{v}_{\mu^2} \in V(\mu)$.

Averaged (C)

Also, an averaged version of Chowla conjecture is present in the literature. As we will see later, this form of averaging will play a special role from the point of view of Sarnak's program, under the name of convergence on short intervals. Matomäki, Radziwiłł, and Tao showed the following result on the order two correlations of μ :

Theorem 4.11 (Matomäki et al. 2015) *We have*

$$\lim_{\substack{M, H \rightarrow \infty \\ \text{with } H=o(M)}} \frac{1}{HM} \sum_{h \leq H} \left| \sum_{m \leq M} \mu(m) \mu(m+h) \right| = 0.$$

As a consequence of the above, we have (cf. Corollary 4.7):

Corollary 4.12 (Ferenczi et al. 2018)

For each Furstenberg system $(S, X_\mu, \mathcal{B}(X_\mu), \kappa)$, the spectral measure σ_F of F is continuous.

(C) vs. Other Conjectures

Chowla conjecture is thought of as a multiplicative analogue of the twin primes problem. Indeed, Twin Primes Conjecture in its quantitative form expects that $\sum_{n \leq N} \Lambda(n) \Lambda(n+2) = \left(2 \prod_{p \geq 3} \left(1 - \frac{1}{(p-1)^2} \right) \right) N + o(N)$, where Λ is the von Mangoldt function: $\Lambda(p^k) = \log p$ for each prime p and $k \geq 1$, and Λ vanishes at all other values of n . This function is a good approximation of $1_{\mathbb{P}}$ and it is **not** a multiplicative function.

Moreover, Chowla conjecture is a special instance of Elliott conjecture on correlations of multiplicative functions. Let $u_0, \dots, u_k : \mathbb{N} \rightarrow \mathbb{U}$ be multiplicative. Elliott conjecture, in a corrected form given in (Matomäki et al. 2015), asserts that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n \leq N} u_0(n+k_0) u_1(n+k_1) \dots u_k(n+k_r) = 0,$$

if for some $0 \leq j \leq k$, u_j is strongly aperiodic. This conjecture was stated first in Elliott (1992) and Elliott (1994), and in its original version turned out to be false, see (Matomäki et al. 2015) for details. Also, a logarithmically averaged version of Elliott conjecture appears in the literature. For the details, see Tao and Teräväinen (2019a, b) and the references therein.

Sarnak's Conjecture

(S)

Sarnak's Conjecture from 2010 states that

each (topological) zero entropy system

(T, X) is Möbius orthogonal, that is, (S)

satisfies (3) for all $f \in C(X)$ and $x \in X$.

As zero entropy expresses the fact that the system is “deterministic” (or of “low complexity”), Sarnak's conjecture captures our expectation that prime numbers behave globally as a random sequence, or, more precisely, that they cannot be predicted by a low-complexity object. One can relax the entropy assumptions on T in Sarnak's conjecture in the following way:

Theorem 5.1 (El Abdalaoui et al. 2017a) (S) *is equivalent to Möbius orthogonality for each topological dynamical system (T, X) , each continuous function $f \in C(X)$ and each completely deterministic point $x \in X$ (i.e. (3) holds at x for all $f \in C(X)$).*

Remark 5.2 The difficulty in Sarnak's conjecture comes from the requirement “for all $x \in X$.” An a.e. version of Möbius orthogonality is true for **all** dynamical systems. The proof makes use of Davenport's estimate (7) (below), see El Abdalaoui et al. (2017a); Sarnak (n.d.).

Even though Sarnak's Conjecture is defined in terms of topological dynamics, it can be translated to ergodic-theoretic language. Namely

$$\frac{1}{N} \sum_{n \leq N} f(T^n x) \mu(n) = \int_{X \times X_\mu} f \otimes Fd \left(\frac{1}{N} \sum_{n \leq N} \delta_{(T^n x, S^n \mu)} \right).$$

Thus, we need to study the properties of joinings given by the limit points of

$$\frac{1}{N} \sum_{n \leq N} \delta_{(T^n x, S^n \mu)}.$$

The simplest case where (S) is known to hold is the one-point dynamical system: $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n \leq N} \mu(n) = 0$ is equivalent to the Prime Number Theorem. (S) for rotations on finite groups is equivalent to Dirichlet's Prime Number Theorem. For irrational rotations, (S) follows from an old (quantitative) result of Davenport (Davenport 1937): for an arbitrary $A > 0$,

$$\max_{t \in \mathbb{T}} \left| \sum_{n \leq N} e^{2\pi i n t} \mu(n) \right| \leq C_A \frac{N}{\log^A N} \quad (7)$$

for some $C_A > 0$ and all $N \geq 2$.

As we will see later, an important role in the research around Sarnak's conjecture is played by nil-systems. Green and Tao obtained the following quantitative version of (S):

Theorem 5.3 (Green and Tao 2012) *Let G be a simply-connected nilpotent Lie group with a discrete and cocompact subgroup Γ . Let $p : \mathbb{Z} \rightarrow G$ be any its polynomial sequence: $p(n) = d_1^{p_1(n)} \dots d_k^{p_k(n)}$, where $p_j : \mathbb{N} \rightarrow \mathbb{N}$ is a polynomial, $j = 1, \dots, k$ and $f : G/\Gamma \rightarrow \mathbb{R}$ a Lipschitz function. Then*

$$\left| \sum_{n \leq N} f(p(n)\Gamma) \mu(n) \right| = O_{f,G,\Gamma,A} \left(\frac{N}{\log^A N} \right)$$

for all $A > 0$.

In particular, all nilrotations are Möbius orthogonal.

(S) vs. (C)

Sarnak's conjecture was originally mainly motivated by Chowla conjecture; we have the following result:

Theorem 5.4 (C) implies (S).

Theorem 5.4 is already stated in (Sarnak n.d.). In fact, it is a purely ergodic theory claim: we have already noticed that both conjectures have their ergodic theory reformulation and a joining proof of Theorem 5.4 can be found in El Abdalaoui et al. (2017a). The main idea is the following: suppose that $\frac{1}{N_k} \sum_{n \leq N_k} \delta_{(T^n x, S^n \mu)} \rightarrow \rho$. The projection of this joining onto X is a zero entropy measure κ , whereas the projection onto X_μ equals $\widehat{v}_S$ by Chowla conjecture. Moreover, $(S, X_\mu, \widehat{v}_S)$ has the property of being relative Kolmogorov with respect to its factor (S, X_{μ^2}, v_S) . On the other hand, the restriction of ρ to $X \times X_{\mu^2}$ is of relative zero entropy over X_{μ^2} . This yields relative disjointness of systems $(S, X_\mu, \widehat{v}_S)$ and $(T \times S, X \times X_{\mu^2}, \rho|_{X \times X_{\mu^2}})$ over their common factor (S, X_{μ^2}, v_S) . To complete the proof, we use the orthogonality of F to $L^2(X_{\mu^2}, v_S)$.

Remark 5.5 It still remains open whether (S) implies (C), see however Remark 5.16.

In Huang et al. (2019a), Möbius orthogonality for low complexity systems is discussed. Following (Ferenczi 1997), we say that the measure-complexity of $\mu \in M(T, X)$ is weaker than $a = (a_n)_{n \geq 1}$ if

$$\liminf_{n \rightarrow \infty} \frac{\min \left\{ m \geq 1 : \mu \left(\bigcup_{j=1}^m B_{d_n}(x_i, \varepsilon) \right) > 1 - \varepsilon \text{ for some } x_1, \dots, x_m \in X \right\}}{a_n} = 0$$

for each $\varepsilon > 0$ (here $d_n(y, z) = \frac{1}{n} \sum_{j=1}^n d(T^j y, T^j z)$).

Theorem 5.6 (Huang et al., 2019a) *Suppose that (C) holds for correlations of order 2 (i.e., for $r = 1$). Then (T, X) is Möbius orthogonal whenever all*

invariant measures for (T, X) are of complexity weaker than n .

To obtain a non-conditional result, Huang, Wang, and Ye used a difficult estimate of Matomäki, Radziwiłł, and Tao (namely, “Truncated Elliott on the average,” applied to μ) from (Matomäki et al. 2015). The cost to be paid is a further strengthening of the assumptions on the complexity of (T, X) .

Theorem 5.7 (Huang et al. 2019a) *Suppose that all invariant measures of (T, X) are of sub-polynomial complexity, that is, their complexity is weaker than $(n^\tau)_{n \geq 1}$ for each $\tau > 0$. Then (T, X) is Möbius orthogonal.*

See Huang et al. (n.d.) for the most recent application of this result.

Finally, let us point out a consequence of the result on correlations of μ of order 2. Directly from Corollary 4.12, we have:

Corollary 5.8 All topological dynamical systems whose all invariant measures yield systems with discrete spectrum are Möbius orthogonal.

In the uniquely ergodic case, an earlier and independent proof of this fact was given by Huang, Wang, and Zhang (Huang et al. 2019b) (for the totally uniquely ergodic case, see El Abdalaoui et al. (2017b)). The result also follows from Huang et al. (2019a).

Strong MOMO Property

Given an arithmetic function u , following El Abdalaoui et al. (2018), we say that (T, X) satisfies the *strong u -OMO property* if, for any increasing sequence of integers $0 = b_0 < b_1 < b_2 < \dots$ with $b_{k+1} - b_k \rightarrow \infty$, for any sequence (x_k) of points in X , and any $f \in C(X)$, we have

$$\frac{1}{b_K} \sum_{k < K} \left| \sum_{b_k \leq n < b_{k+1}} f(T^{n-b_k} x_k) u(n) \right| \xrightarrow{K \rightarrow \infty} 0. \quad (8)$$

If $u = \mu$, we speak about the strong MOMO property. (The acronym MOMO stands for Möbius Orthogonality of Moving Orbits.) Strong

MOMO property was introduced in El Abdalaoui et al. (2018) to deal with Möbius orthogonality of uniquely ergodic models of a given measure-theoretic dynamical system. Moreover, we have:

Theorem 5.9 (El Abdalaoui et al. 2018) *The following conditions are equivalent:*

- (i) *All zero entropy systems are Möbius orthogonal, that is, Sarnak's conjecture holds.*
- (ii) *For each zero entropy system (T, X) , we have $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n \leq N} f(T^n x) \mu(n) = 0$ when $N \rightarrow \infty$, for each $f \in C(X)$, uniformly in $x \in X$, that is, uniform Sarnak's conjecture holds.*
- (iii) *All zero entropy systems enjoy the strong MOMO property.*

By taking $f = 1$, we obtain that strong u -OMO implies the following:

$$\frac{1}{b_K} \sum_{k < K} \left| \sum_{b_k \leq n < b_{k+1}} u(n) \right| \xrightarrow{K \rightarrow \infty} 0 \quad (9)$$

for every sequence $0 = b_0 < b_1 < b_2 < \dots$ with $b_{k+1} - b_k \rightarrow \infty$. In particular, $\frac{1}{N} \sum_{n \leq N} u(n) \xrightarrow{N \rightarrow \infty} 0$.

In a similar way (by considering finite rotations), one can deduce $\frac{1}{N} \sum_{n \leq N} u(an + b) \xrightarrow{N \rightarrow \infty} 0$. Thus, (8) can be seen as a form of aperiodicity. A further analysis reveals that, in fact, we deal with a special behaviour of u on a typical short interval. All strongly aperiodic multiplicative functions satisfy (9) (this follows from Theorem A.1 (Matomäki et al. 2015)), hence condition (9) is satisfied both for μ and λ , cf. section “Behaviour on Short Intervals.”

Recently, in Gornilko et al. (2020), the strong u -OMO property was rephrased in the language of functional analysis; and it is equivalent to

$$\begin{aligned} \lim_{K \rightarrow \infty} \frac{1}{b_{K+1}} \sum_{k \leq K} \left\| \sum_{b_k \leq n < b_{k+1}} u(n) f \circ T^n \right\|_{C(X)} \\ = 0 \text{ for all } f \in C(X), (b_k) \text{ as above.} \end{aligned}$$

Usefulness of the strong MOMO concept is seen in the following result:

Proposition 5.10 (El Abdalaoui et al. 2018) *If (R, Z, D, κ) is an ergodic (measure-theoretic) dynamical system and (T, X) is its uniquely ergodic model satisfying the strong MOMO property, then all uniquely ergodic models of (R, Z, D, κ) are Möbius orthogonal. In fact, the strong MOMO holds in all of them.*

Möbius orthogonality of Positive Entropy Systems

If we take a positive entropy system (T, X) , it is natural to expect that it is not Möbius orthogonal. Indeed, trivially, the full shift on $\{0, 1\}^{\mathbb{Z}}$ is not, and more generally subshifts of finite type are not, see Karagulyan (2017). One can also show that the subshift (S, X_{μ^2}) (which is of positive entropy, see Peckner (2015)) is not Möbius orthogonal, despite the fact that μ^2 itself is a completely deterministic point and Möbius orthogonality holds at it: $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n \leq N} f(S^n \mu^2) \mu(n) = 0$ for each $f \in C(X_{\mu^2})$, see Ferenczi et al. (2018). However, Sarnak's conjecture does not exclude a possibility that **some** positive entropy system is also Möbius orthogonal. (It is mentioned in Sarnak (n.d.) that Bourgain (unpublished) had such a construction.) Downarowicz and Serafin proved the following general result:

Theorem 5.11 (Downarowicz and Serafin 2019a) *Fix an integer $N \geq 2$. Let $\mathbf{u}$ be any bounded, real, aperiodic sequence. Then, there exists a subshift (S, X) over N symbols of entropy arbitrarily close to $\log N$, uncorrelated to $\mathbf{u}$: $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n \leq N} f(S^n x) \mathbf{u}(n) = 0$ for each $f \in C(X)$ and $x \in X$.*

Even more surprisingly, they proved a uniform version of the above result:

Theorem 5.12 (Downarowicz and Serafin 2019b) *Under the same assumption on $\mathbf{u}$, given $N \geq 2$, there exists a strictly ergodic subshift over N symbols, of entropy arbitrarily close to $\log N$, uniformly uncorrelated to $\mathbf{u}$.*

Realizing that, one might be anxious what finally is the class of systems which are Möbius

orthogonal, and in particular, why zero entropy should play a special role. As we will see, however, positive entropy systems are not expected to enjoy the strong MOMO property, cf. Theorem 5.9. Indeed, the following has been proved in El Abdalaoui et al. (2018) (the result has been proved in El Abdalaoui et al. (2018) for the Liouville function but it can also be proved for μ):

Theorem 5.13 (El Abdalaoui et al. 2018) *Let $u \in \{-1, 0, 1\}^{\mathbb{Z}}$ be a generic point for the measure $\hat{\nu}_{\mu^2}$. Then the following conditions are equivalent:*

- (i) (T, X) satisfies strong $\mathbf{u}$ -OMO property.
- (ii) (T, X) is of zero entropy.

As an immediate consequence, we have the following:

Corollary 5.14 *If Chowla conjecture holds, then a system (T, X) has the strong MOMO property if and only if it has zero entropy.*

($S_{\log}$) vs. ($C_{\log}$)

The logarithmic version of Sarnak's conjecture was formulated in Tao (2016) along with ($C_{\log}$) and it postulates that

$$\lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{n \geq N} \frac{1}{n} f(T^n x) \mu(n) = 0 \quad (S_{\log})$$

(with all parameters as in (4)). In Tao (2017), Tao showed the following:

Theorem 5.15 $(S_{\log})$ is equivalent to $(C_{\log})$.

Remark 5.16 Combining Theorem 5.15 with Corollary 4.3, we obtain that $(S_{\log})$ implies (C) along a subsequence of logarithmic density 1. In particular, (S) implies (C) along a subsequence of full logarithmic density.

Let us here recall one more “logarithmic conjecture” from (Tao 2017) which confirms a special role played by nil-systems in dynamics. Let $(T_{g_0}, G/\Gamma)$ be a nilrotation. Let $f \in C(G/\Gamma)$ be

Lipschitz continuous and $x_0 \in G$. Then (for $H \leq N$ with $H \rightarrow \infty$)

$$\sum_{n \leq N} \frac{\sup_{g \in G} \left| \sum_{h \leq H} f \left(T_g^{h+n} (x_0 \Gamma) \right) \mu(n+h) \right|}{n} = o(H \log N). \quad \left(\mathbf{S}_{\log}^{\text{nil}} \right)$$

Theorem 5.17 (Tao 2017) $\left(\mathbf{S}_{\log}^{\text{nil}} \right)$ is equivalent to $(\mathbf{S}_{\log})$ (and $(\mathbf{C}_{\log})$).

Finally, as a consequence of the result on logarithmic correlations of μ of order 2 (using Corollary 4.7), we obtain:

Corollary 5.18 All topological dynamical systems whose all invariant measures yield systems with singular spectrum are logarithmically Möbius orthogonal.

In general, we do not know if we can replace “all invariant measures” with “all ergodic invariant measures” in the above corollary (the same applies to Corollary 5.8). This replacement is possible, however, when there are only countably many ergodic invariant measures, cf. the discussion on Frantzikinakis and Host (2018) in section “[Logarithmic Furstenberg Systems](#).”

(S) vs. $(\mathbf{S}_{\log})$

Clearly, (S) implies $(\mathbf{S}_{\log})$. As for the other direction, we have the following:

Theorem 5.19 (Gomilko et al. 2020) *Suppose that $(\mathbf{S}_{\log})$ holds. Then there exists a sequence of logarithmic density 1, along which (S) holds for all zero entropy topological dynamical systems.*

The idea of the proof is to use Theorem 5.15, Remark 5.16, and then repeat the arguments from the proof of Theorem 5.4.

Notice that the sequence of logarithmic density 1 in the above result is universal for all zero entropy systems. In Gomilko et al. (2020), one more version of Möbius orthogonality is studied,

namely so-called *logarithmic strong MOMO property* (cf. section “[Strong MOMO Property](#)”):

$$\lim_{K \rightarrow \infty} \frac{1}{\log b_{K+1}} \sum_{k \leq K} \left\| \sum_{b_k \leq n \leq b_{k+1}} \frac{\mu(n)}{n} f \circ T^n \right\|_{C(X)} = 0.$$

Equivalently, for all increasing sequences $(b_k) \subset \mathbb{N}$ with $b_{k+1} - b_k \rightarrow \infty$, all $(x_k) \subset X$ and $f \in C(X)$,

$$\lim_{K \rightarrow \infty} \frac{1}{\log b_{K+1}} \sum_{k \leq K} \left| \sum_{b_k \leq n < b_{k+1}} \frac{1}{n} f(T^{n-b_k} x_k) \mu(n) \right| = 0.$$

Theorem 5.20 (Gomilko et al. 2020) *Assume that a topological system (T, X) satisfies the logarithmic strong MOMO property. Then there exists a sequence $A = A(T, X) \subset \mathbb{N}$ with full logarithmic density such that, for each $f \in C(X)$,*

$$\lim_{A \ni N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n \leq N} \mu(n) f \circ T^n \right\|_{C(X)} = 0.$$

In particular, Möbius orthogonality holds along a subsequence of full logarithmic density.

Remark 5.21 In Gomilko et al. (2020), using Frantzikinakis and Host (2018), it is proved that each system (T, X) for which $M^e(T, X)$ is countable satisfies the logarithmic strong MOMO property, hence, for each such system Sarnak’s conjecture holds in (logarithmic) density, cf. Theorem 6.10.

Strategies

The first years of activity around Sarnak’s conjecture were devoted to proving Möbius orthogonality in **selected classes** of zero entropy dynamical systems. While this proved fruitful, and often some brilliant arguments were found ad hoc, with a strong dependence on the class under consideration, it quickly became clear that it won’t be sufficient. Two main strategies to attack (S) arose:

A. The first strategy is to look for some additional intrinsic structure in zero entropy systems that

could be used to prove orthogonality from μ , namely, **internal disjointness**. Here, a priori, one does not use any other property of μ than multiplicativity and boundedness.

- B. As we have seen, (S) is intimately related to (C), and therefore one cannot expect to confirm (S) without using further **number-theoretic properties of μ** . This directs attention to **aperiodicity** and behavior on the so-called **short intervals**. It extends further to studying **Furstenberg systems of μ** (including the logarithmic ones) and trying to interpret arithmetic properties of μ as ergodic properties of the corresponding dynamical systems. One can hope finally to deduce (some kind of) Furstenberg (!) disjointness of Furstenberg systems of μ with a wide subclass of zero entropy systems (hopefully, with all such systems).

As we will see, these two approaches often intertwine, proving once again that number theory and ergodic theory should not be studied separately from each other.

Arithmetic Properties of the Möbius Function

Multiplicativity

Internal disjointness. Joinings (introduced in a seminal paper of Furstenberg (Furstenberg 1967)) have been present in ergodic theory for over 50 years. Disjointness (absence of nontrivial joinings), as a form of an extremal non-isomorphism and a measure-theoretic invariant, has always played a crucial role in classification problems. (Recall also that different powers for a typical automorphism of a standard Borel space are pairwise disjoint (del Junco 1981). See also more recent Kanigowski et al. (2020).) It appeared, however, in many other contexts, including homogenous dynamics, with applications in number theory. Sarnak's conjecture gave yet a new impetus, in particular for studying (approximate) disjointness for different sub-actions.

A basic method to prove orthogonality with a multiplicative function comes from the multiplicative orthogonality criterion (MOC):

Theorem 6.1 (Bourgain et al. 2013; Daboussi, Kátai 1986) *Assume that $(f_n) \subset \mathbb{C}$ is a bounded sequence. Assume that for all (sufficiently large) prime numbers $p \neq q$,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n \leq N} f_{pn} \bar{f}_{qn} = 0. \quad (10)$$

Then, for each bounded multiplicative function u , we have $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n \leq N} f_n u(n) = 0$. In particular, (f_n) is Möbius orthogonal.

Remark 6.2 Notice that Theorem 6.1 does not require from u anything but multiplicativity and boundedness.

In the dynamical context (T, X) the simplest way to use Theorem 6.1 is to take $f_n = f(T^n x)$. In this form, MOC appeared for the first time in Bourgain et al. (2013) and was used to prove that the horocycle flows are Möbius orthogonal. To see how MOC is used and how it is related to Furstenberg's disjointness theory (Furstenberg 1967), assume that $M(T, X) = \{\mu\}$, $\int_X f d\mu = 0$, and the corresponding measure-theoretic system is totally ergodic. Then, any measure $\rho \in \mathcal{V}(x, x)$ (considered in the topological dynamical system $(T^p \times T^q, X \times X)$) is a joining of T^p and T^q . If we now assume that (T^p, X, μ) and (T^q, X, μ) are disjoint for sufficiently large primes $p \neq q$, then $\rho = \mu \otimes \mu$ and, as a result, the limit in (10) equals $\int_{X \times X} f \otimes \bar{f} d\rho = 0$, that is, the assumptions of MOC are satisfied.

In general, a use of MOC is not that simple. Consider an irrational rotation $Tx = x + \alpha$ on the circle $X = \mathbb{R}/\mathbb{Z}$. To see that (10) holds for all characters, one uses the Weyl criterion on uniform distribution. However, there are continuous zero mean functions for which (10) fails (Kułaga-Przymus and Lemańczyk 2015), which shows clearly, that in general we can only expect (10) to hold for a linearly dense set of continuous functions.

In some cases, MOC cannot be applied directly (e.g., when the systems under consideration fail to be weakly mixing) and the spectral approach can help. Examples can be found in (El Abdalaoui et al. (2014), (2016), and Bourgain (2013)).

AOP property. The following ergodic counterpart of MOC was developed in El Abdalaoui et al. (2017b): an ergodic automorphism $(T, X, \mathcal{B}, \mu)$ is said to have *asymptotically orthogonal powers* (AOP) if for each $f, g \in L_0^2(X, \mathcal{B}, \mu)$, we have

$$\lim_{\substack{p \rightarrow \infty, q \rightarrow \infty, p \neq q}} \sup_{\kappa \in J^c(T^p, T^q)} \left| \int_{X \times X} f \otimes g \, d\kappa \right| = 0. \quad (11)$$

Clearly, if the powers of T are pairwise disjoint, then T enjoys the AOP property. However, this condition is not necessary, the powers of T having AOP property may even be isomorphic. Moreover, AOP implies total ergodicity and zero entropy (El Abdalaoui et al., 2017b). The relation between strong MOMO and AOP properties is described by the following result:

Theorem 6.3 (El Abdalaoui et al. 2017a, b) *Let $\mathbf{u}$ be a bounded multiplicative function. Suppose that $(R, \mathcal{Z}, \mathcal{C}, \kappa)$ satisfies AOP. Then the following are equivalent:*

- $\mathbf{u}$ satisfies (9);
- The strong $\mathbf{u}$ -OMO property is satisfied in each uniquely ergodic model (T, X) of R .

In particular, if the above holds, for each $f \in C(X)$, we have

$$\frac{1}{N} \sum_{n \leq N} f(T^n x) \mathbf{u}(n) \xrightarrow{N \rightarrow \infty} 0 \text{ uniformly on } X.$$

Aperiodicity

As all periodic sequences are orthogonal to μ , one can expect that sequences with some properties similar to periodicity will also be Möbius orthogonal. Notice also that Möbius orthogonality of periodic sequences ((S) for rotations on finite groups) corresponds to μ being aperiodic. This is the simplest situation where some additional

properties of μ (other than multiplicativity) begin to play a significant role.

Zero entropy continuous interval maps. In Karagulyan (2015), (S) for zero entropy continuous intervals maps and orientation-preserving circle homeomorphisms is established. The starting point for developing the main tools is the result of Davenport (7), which shows clearly that the examples under consideration are indeed “relatives” of irrational rotations. Additionally, in order to treat the case of interval maps one studies ω -limit sets and it turns out that, in fact, one deals with an odometer.

Synchronized automata. In Deshouillers et al. (2015), Deshouillers, Drmota, and Müllner prove that (S) is true for automatic sequences generated by synchronizing automata (the inputs are read with the most significant digit first). In fact, they prove orthogonality of such sequences from any bounded function $\mathbf{u}$ that is aperiodic.

Almost periodic sequences. We say that a sequence is *Weyl rationally almost periodic* (WRAP) whenever it can be approximated arbitrarily well by periodic sequences in Weyl pseudo-metric d_W given by $d_W(x, y) = \limsup_{N \rightarrow \infty} \sup_{l \geq 1} \frac{1}{N} |\{l \leq n \leq l + N : x(n) \neq y(n)\}|$. It is proved in Bergelson et al. (2019) that each subshift (S, X_s) given by a Weyl almost periodic sequence is Möbius orthogonal (in fact, we have orthogonality to any bounded aperiodic arithmetic function $\mathbf{u}$).

Behaviour on Short Intervals

During the last four years, an enormous progress concerning the short interval behavior of strongly aperiodic multiplicative functions has been made due to the breakthrough result of Matomäki and Radziwiłł. The main result of Matomäki and Radziwiłł (2016), for μ , in its simplified form can be written as

$$\lim_{\substack{M, H \rightarrow \infty \\ \text{with } H=o(M)}} \frac{1}{M} \sum_{1 \leq m \leq M} \frac{1}{H} \left| \sum_{m \leq h < m+H} \mu(h) \right| = 0. \quad (12)$$

This gave an impetus to study convergence on short intervals in ergodic theory and it has become

a new, crucial player from the point of view of Sarnak's conjecture. Condition (12) can be also reformulated in the following way: for each $(b_n) \subset \mathbb{N}$ with $b_{n+1} - b_n \rightarrow \infty$, we have

$$\lim_{K \rightarrow \infty} \frac{1}{b_{K+1}} \sum_{k \leq K} \left| \sum_{b_k \leq n < b_{k+1}} \mu(n) \right| = 0, \quad (13)$$

cf. Section "Strong MOMO Property" 5.3.

Almost periodic sequences. In Bergelson et al. (2019), in case of WRAP x , the authors also ask about the behavior of averages of the form

$$\frac{1}{H} \sum_{m \leq h < m+H} f(S^h z) \mu(n) \quad (14)$$

(where $z \in X_x$) for large values of H and arbitrary $m \in \mathbb{N}$. Under (C), convergence to zero uniformly in m does not take place; however, it is shown in Bergelson et al. (2019) that for a "typical" $m \in \mathbb{N}$ the averages in (14) are small. The key argument in the proof comes from a result of Matomäki, Radziwiłł, and Tao:

Theorem 6.4 (Matomäki et al. 2015) *For each periodic sequence $a(n)$, we have*

$$\lim_{\substack{H, M \rightarrow \infty \\ H=o(M)}} \frac{1}{M} \sum_{M \leq m < 2M} \left| \frac{1}{H} \sum_{m \leq h < m+H} a(h) \mu(h) \right| = 0. \quad (15)$$

As a consequence, we have:

Theorem 6.5 (Bergelson et al. 2019) *Suppose that $x \in \mathbb{A}^{\mathbb{Z}}$ is WRAP. Then for all $f \in C(X_x)$ and $z \in X_x$,*

$$\lim_{\substack{H, M \rightarrow \infty \\ H=o(M)}} \frac{1}{M} \sum_{M \leq m < 2M} \left| \frac{1}{H} \sum_{m \leq h < m+H} f(S^h z) \mu(h) \right| = 0. \quad (16)$$

Moreover, it is shown that all synchronized automata yield WRAP sequences. Thus, the above theorem strengthens the aforementioned result by Deshouillers, Drmota, and Müllner in Deshouillers et al. (2015).

Rigid systems. In Kanigowski et al. (n.d.-a), Kanigowski, Lemańczyk, and Radziwiłł study rigid systems. (A measure-theoretic system $(R, Z, \mathcal{C}, \kappa)$ is rigid if, for some increasing sequence (q_n) of natural numbers, we have $f \circ R^{q_n} \rightarrow f$ in $L^2(Z, \kappa)$ for each $f \in L^2(Z, \kappa)$. Rigid systems are of zero entropy. Moreover, the typical measure-theoretic automorphism is rigid and weakly mixing.) To formulate their results, we need some definitions and facts. Given a natural number q , the sum $\sum_{p|q} 1/p$ is called the *prime volume* of q . The prime volume grows slowly with q :

$$\sum_{p|q} \frac{1}{p} \leq \log \log \log q + O(1).$$

However "most" of the time, the prime volume of q stays bounded: if we set

$$\mathcal{D}_j := \left\{ q \in \mathbb{N} : \sum_{p|q} \frac{1}{p} < j \right\},$$

then $d(\mathcal{D}_j) \rightarrow 1$ when $j \rightarrow \infty$. A topological system (T, X) is said to be *good* if for every $v \in M(T, X)$ at least one of the following conditions holds:

- **(BPV rigidity):** $(T, X, \mathcal{B}, v)$ is rigid along a sequence $(q_n)_{n \geq 1}$ with **bounded prime volume**, that is, there exists j such that $(q_n)_{n \geq 1} \subset \mathcal{D}_j$;
- **(PR rigidity):** $(T, X, \mathcal{B}, v)$ has **polynomial rate** of rigidity, that is, there exists a linearly dense (in $C(X)$) set $\mathcal{F} \subset C(X)$ such that for each $f \in \mathcal{F}$ we can find $\delta > 0$ and a sequence $(q_n)_{n \geq 1}$ satisfying

$$\sum_{j=-q_n^\delta}^{q_n^\delta} \|f \circ T^{jq_n} - f\|_{L^2(v)}^2 \rightarrow 0.$$

They prove the following:

Theorem 6.6 (a) Assume the (T, X) is a good topological system. Then (T, X) is Möbius orthogonal.

(b) Suppose that each ergodic invariant measure of (T, X) yields either BPV rigidity or PR rigidity and $M^e(T, X)$ is countable then (T, X) is Möbius orthogonal.

A key tool here is a strengthening of the main result of Matomäki and Radziwiłł (Matomäki & Radziwiłł 2016) (cf. (12)) to short interval behaviour along arithmetic progressions:

Theorem 6.7 (Kanigowski et al. n.d.-a) For each $\varepsilon > 0$, there exists L_0 such that for each $L \geq L_0$ and $q \geq 1$ satisfying $\sum_{p|q} 1/p \leq (1 - \varepsilon) \sum_p \leq L/1/p$ we can find $M_0 = M_0(q, L)$ such that for all $M \geq M_0$, we have

$$\sum_{j=0}^{M/Lq} \sum_{a=0}^{q-1} \left| \sum_{\substack{m \in [z + jLq, z + (j+1)Lq] \\ m \equiv a \pmod{q}}} \mu(m) \right| < \varepsilon M$$

for some $0 \leq z < Lq$.

Remark 6.8 Despite the fact that PR rigidity does not seem to be stable under different (uniquely ergodic) models of a measure-preserving transformation, assuming (T, X) is uniquely ergodic, it is proved that (T, X) satisfies the strong MOMO property whenever its unique invariant measure yields either BPV or PR rigidity. Via Proposition 5.10, we obtain that if in a model PR rigidity holds, all of the models are Möbius orthogonal. This, in particular, applies to all ergodic transformations with discrete spectrum. Moreover, it is shown in Kanigowski et al. (n.d.-a) that for a.e. IET (of $d \geq 3$ intervals) BPV rigidity holds, so a.e. IET (and all their uniquely ergodic models) is Möbius orthogonal. This is to be compared with previously known results for 3-IETs (Bourgain 2013; Chaika and Eskin 2019;

Ferenczi and Mauduit 2018; Karagulyan 2020). Other applications are given for $C^{2+\varepsilon}$ Anzai skew products and for some so-called Rokhlin extensions of rotations.

One more of the consequences is the following result:

Corollary 6.9 (Kanigowski et al. n.d.-a) No Furstenberg system of the Möbius function μ is either BPV or PR rigid. The same holds for the Liouville function λ .

Logarithmic Furstenberg Systems

Frantzikinakis and Host study logarithmic Furstenberg systems associated to μ (and λ). They prove the following remarkable result:

Theorem 6.10 (Frantzikinakis and Host 2018) Each zero entropy topological system (T, X) with only countably many ergodic measures is logarithmically Möbius orthogonal.

In particular, uniquely ergodic systems of zero topological entropy satisfy $(S_{\log})$. The key argument in the proof of Theorem 6.10 is the following structural result on the logarithmic Furstenberg systems of μ and λ :

Theorem 6.11 (Frantzikinakis and Host 2018) Each logarithmic Furstenberg system of μ or λ is a factor of a system that:

- has no irrational spectrum,
- has ergodic components isomorphic to direct products of infinite-step nilsystems and Bernoulli systems.

The starting point for the proof of the above theorem, resulting in a reduction of the problem to purely ergodic context, is an identity of Tao (implicit in (Tao 2016)) showing that self-correlations of μ (and λ) are averages of its dilated self-correlations with prime dilates. Frantzikinakis and Host also prove that logarithmic Furstenberg systems of μ (and λ) are “almost

determined” by strongly stationary processes (introduced by Furstenberg and Katznelson in the 1990s). The structure of (measure-theoretic) dynamical systems given by strongly stationary processes has been described by Jenvey (Jenvey 1997) who proved that ergodicity implies Bernoulli (cf. Theorems 4.2 and 4.4) and Frantzikinakis (Frantzikinakis 2004) who described the ergodic decomposition in the non-ergodic case.

The above results are extended in Frantzikinakis and Host (2019) to strongly aperiodic multiplicative functions. Moreover, the following multidimensional result is proved:

Theorem 6.12 *Let $f_1, \dots, f_l: \mathbb{N} \rightarrow \mathbb{U}$ be multiplicative functions. Let (R, Y) be a topological dynamical system and let $y \in Y$ be a logarithmically generic point for a measure ν with zero entropy and having at most countably many ergodic components, all of which are totally ergodic. Then for every $g \in C(Y)$ that is orthogonal in $L^2(\nu)$ to all R -invariant functions we have*

$$\lim_{N \rightarrow \infty} \frac{1}{\log N} \sum_{n \leq N} \frac{g(R^n y) \prod_{j=1}^l f_j(n + h_j)}{n} = 0 \quad (17)$$

for all $h_1, \dots, h_l \in \mathbb{Z}$.

The unweighted version of (with $g = 1$) is expected to hold if the shifts are distinct and at least one of the multiplicative functions is strongly aperiodic. This is the logarithmically averaged version of Elliott conjecture (Elliott 1990, 1994; Matomäki et al. 2015). In the special case of irrational rotations, with $l = 1$, Theorem 6.12 is the logarithmically averaged variant of a classical result of Daboussi (Collectif 1975; Daboussi and Delange 1974, 1982). Already in case $l = 2$ it is completely new.

Future Directions

Detecting Zero Entropy

As shown in Theorem 5.13 and Corollary 5.14, under the Chowla conjecture, the Möbius function is a sequence that:

- is strong MOMO orthogonal to all zero entropy systems,
- is never strong MOMO orthogonal to any positive entropy system.

In view of this, it is natural to study the following problem:

Question 1 Which numerical sequences distinguish between zero and positive entropy systems?

Note that in view of the results of Downarowicz and Serafin (Theorems 5.11 and 5.12), “usual orthogonality” or even its uniform version is insufficient for these needs.

Proving the Strong MOMO Property

It was already asked in Ferenczi et al. (2018) whether whenever we have a zero entropy system (T, X) for which we can prove Möbius orthogonality, then we can prove the strong MOMO property. Recently, Lemńczyk and Müllner (n.d.) proved the strong MOMO property for (primitive) automatic sequences (previously known to be Möbius orthogonal by Müllner (2017)), answering a question from Ferenczi et al. (2018). Yet another question persists:

Question 2 Do horocycle flows satisfy the strong MOMO property?

We do not even know whether Möbius orthogonality takes place in all uniquely ergodic models of horocycle flows.

Mixing Properties of Furstenberg Systems

Corollary 6.9 induces the following problem:

Question 3 Are Furstenberg systems of λ mildly mixing?

For μ we need to take into account that its Furstenberg systems have the discrete spectrum factor given by the Mirsky measure of μ^2 .

Furstenberg Disjointness in Non-ergodic Case

As Host and Frantzikinakis' analysis shows, if the (potential) logarithmic Furstenberg systems of λ or μ are non-ergodic, then they are very non-ergodic. One of open questions by Frantzikinakis is whether the system $\mathbb{T}^2 \ni (x, y) \mapsto (x, x + y) \in \mathbb{T}^2$ considered with Lebesgue measure can be a Furstenberg system of λ . Of course this example is a measure-theoretic system which is Furstenberg disjoint from all ergodic systems. It seems to be a problem of independent interest to fully understand the class of transformations disjoint from all ergodic transformations.

Bibliography

- Bergelson V, Kułaga-Przymus J, Lemańczyk M, Richter FK (2019) Rationally almost periodic sequences, polynomial multiple recurrence and symbolic dynamics. *Ergodic Theory Dyn Syst* 39(9):2332–2383
- Bourgain J (2013) On the correlation of the Moebius function with rank-one systems. *J. Anal. Math.* 120:105–130
- J. Bourgain, P. Sarnak, and T. Ziegler. Disjointness of Möbius from horocycle flows. In *From Fourier analysis and number theory to Radon transforms and geometry*, volume 28 of *Dev. Math.*, pages 67–83. Springer, New York, 2013.
- Chaika J, Eskin A (2019) Möbius disjointness for interval exchange transformations on three intervals. *Journal of Modern Dynamics* 14(1):55–86
- Chowla S (1965) The Riemann hypothesis and Hilbert's tenth problem. *Mathematics and its applications*, vol 4. Gordon and Breach Science Publishers, New York
- Collectif (1975) Fonctions multiplicatives presque périodiques B. In *Journées arithmétiques de Bordeaux*, number 24–25 in *Astérisque*, pp 321–324. Société mathématique de France
- Daboussi H, Delange H (1974) Quelques propriétés des fonctions multiplicatives de module au plus égal à 1. *C R Acad Sci Paris Sér A* 278:657–660
- Daboussi H, Delange H (1982) On multiplicative arithmetical functions whose modulus does not exceed one. *J Lond Math Soc* 26(2):245–264
- Davenport H (1937) On some infinite series involving arithmetical functions. II. *Quart J Math Oxford* 8:313–320
- del Junco A (1981) Disjointness of measure-preserving transformations, minimal selfjoinings and category. In: *Ergodic theory and dynamical systems, I* (College Park, Md., 1979–80), volume 10 of *Progress in mathematics*, pp 81–89. Birkhäuser, Boston
- Deshouillers J-M, Drmota M, Müllner C (2015) Automatic sequences generated by synchronizing automata fulfill the Sarnak conjecture. *Studia Math* 231(1):83–95
- Downarowicz T (2011) *Entropy in dynamical systems*, volume 18 of *New Mathematical Monographs*. Cambridge University Press, Cambridge
- Downarowicz T, Serafin J (2019a) Almost full entropy subshifts uncorrelated to the Möbius function. *Int Math Res Not* 2019(11):3459–3472
- Downarowicz T, Serafin J (2019b) A strictly ergodic, positive entropy subshift uniformly uncorrelated to the Möbius function. *Studia Math* 251(2020):195–206. Published online (Online First)
- El Abdalaoui EH, Lemańczyk M, de la Rue T (2014) On spectral disjointness of powers for rank-one transformations and Möbius orthogonality. *J Funct Anal* 266(1):284–317
- El Abdalaoui EH, Kasjan S, Lemańczyk M (2016) 0-1 sequences of the Thue-Morse type and Sarnak's conjecture. *Proc Am Math Soc* 144(1):161–176
- El Abdalaoui EH, Kułaga-Przymus J, Lemańczyk M, de la Rue T (2017a) The Chowla and the Sarnak conjectures from ergodic theory point of view. *Discrete Contin Dyn Syst* 37(6):2899–2944
- El Abdalaoui EH, Lemańczyk M, de la Rue T (2017b) Automorphisms with Quasidiscrete spectrum, multiplicative functions and average orthogonality along short intervals. *Int Math Res Not* 2017(14):4350–4368
- El Abdalaoui EH, Kułaga-Przymus J, Lemańczyk M, de la Rue T (2018) Möbius disjointness for models of an ergodic system and beyond. *Israel J Math* 228(2):707–751
- Elliott PDA (1990) Multiplicative functions $|g| \leq 1$ and their convolutions: an overview. In: *Séminaire de Théorie des Nombres, Paris 1987–88*, volume 81 of *Progress in mathematics*, pp 63–75. Birkhäuser Boston, Boston
- Elliott PDA (1992) On the correlation of multiplicative functions. *Notas Soc Mat Chile* 11(1):1–11
- Elliott PDA (1994) On the correlation of multiplicative and the sum of additive arithmetic functions. *Mem Am Math Soc* 112(538):viii+88
- Ferenczi S (1997) Measure-theoretic complexity of ergodic systems. *Isr J Math* 100(1):189–207
- S. Ferenczi, J. Kułaga-Przymus, and M. Lemańczyk. Sarnak's conjecture: what's new. In S. Ferenczi, J. Kułaga-Przymus, and M. Lemańczyk, *Ergodic theory and dynamical systems in their interactions with arithmetics and combinatorics*, volume 2213 of *Lecture Notes in Math.*, 163–235. Springer, Cham, 2018.
- Ferenczi S, Mauduit C (2018) On Sarnak's conjecture and Veech's question for interval exchanges. *J Anal Math* 134(2):545–573

- Frantzikinakis N (2004) The structure of strongly stationary systems. *J Anal Math* 93(1):359–388
- Frantzikinakis N (2017) Ergodicity of the Liouville system implies the Chowla conjecture. *Discrete Anal*, Paper no. 19, 41pp
- Frantzikinakis N, Host B (2018) The logarithmic Sarnak conjecture for ergodic weights. *Ann of Math* 187(3):869–931
- Frantzikinakis N, Host B (2019) Furstenberg systems of bounded multiplicative functions and applications. *Int Math Res Not*. <https://doi.org/10.1093/imrn/rnz037>
- Furstenberg H (1967) Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math Syst Theory* 1:1–49
- Gomilko A, Kwietniak D, Lemańczyk M (2018) Sarnak's conjecture implies the Chowla conjecture along a subsequence. In: Ferenczi S, Kułaga-Przymus J, Lemańczyk M (eds) *Ergodic theory and dynamical systems in their interactions with arithmetics and combinatorics*, volume 2213 of *Lecture Notes in Mathematics*. Springer, Cham, pp 237–247
- Gomilko A, Lemańczyk M, de la Rue T (2020) Möbius orthogonality in density for zero entropy dynamical systems. *Pure and Applied Functional Analysis* 5:1357–1376
- Green B, Tao T (2012) The Möbius function is strongly orthogonal to nilsequences. *Ann Math* 175(2):541–566
- Huang W, Wang Z, Ye X (2019a) Measure complexity and Möbius disjointness. *Adv Math* 347:827–858
- Huang W, Wang Z, Zhang G (2019b) Möbius disjointness for topological models of ergodic systems with discrete spectrum. *J Mod Dyn* 14:277–290
- Huang W, Liu J, Wang K. (n.d.) Möbius disjointness for skew products on a circle and a nilmanifold. Preprint. <https://arxiv.org/abs/1907.01735>
- Jenvey E (1997) Strong stationarity and de Finetti's theorem. *J Anal Math* 73(1):1–18
- Kamae T (1973) Subsequences of normal sequences. *Isr J Math* 16:121–149
- Kanigowski A, Lemańczyk M, Radziwiłł M. (n.d.-a) Rigidity in dynamics and Möbius disjointness, to appear in *Fundamenta Math*. <https://arxiv.org/abs/1905.13256>
- Kanigowski A, Lemańczyk M, Ulcigrai C. (2020) On disjointness properties of some parabolic flows. *Invent Math* 221(1):1–111. <https://arxiv.org/abs/1810.11576>
- Karagulyan D (2015) On Möbius orthogonality for interval maps of zero entropy and orientation-preserving circle homeomorphisms. *Ark Mat* 53(2):317–327
- Karagulyan D (2017) On Möbius orthogonality for subshifts of finite type with positive topological entropy. *Stud Math* 237(3):277–282
- Karagulyan D (2020) Hausdorff dimension of a class of three-interval exchange maps. *Discrete and Contin Dynam Syst* 40(3):1257–1281
- Kátai I (1986) A remark on a theorem of H. Daboussi. *Acta Math Hungar* 47(1–2):223–225
- Kułaga-Przymus J, Lemańczyk M (2015) The Möbius function and continuous extensions of rotations. *Monatsh Math* 178(4):553–582
- Lemańczyk M, Müllner (n.d.) C. Automatic sequences are orthogonal to aperiodic multiplicative functions. Published online in *Discrete Continuous Dynamical Systems* (2020). <https://doi.org/10.3934/dcds.2020260>
- Matomäki K, Radziwiłł M (2016) Multiplicative functions in short intervals. *Ann Math* 183(3):1015–1056
- Matomäki K, Radziwiłł M, Tao T (2015) An averaged form of Chowla's conjecture. *Algebra Number Theory* 9(9):2167–2196
- Müllner C (2017) Automatic sequences fulfill the Sarnak conjecture. *Duke Math J* 166(17):3219–3290
- Peckner R (2015) Uniqueness of the measure of maximal entropy for the squarefree flow. *Isr J Math* 210(1):335–357
- Ram Murty M, Vatwani A (2018) A remark on a conjecture of Chowla. *J Ramanujan Math Soc* 33(2):111–123
- Ramaré O (2018) Chowla's conjecture: from the Liouville function to the Möbius function. In: Ferenczi S, Kułaga-Przymus J, Lemańczyk M (eds) *Ergodic theory and dynamical systems in their interactions with arithmetics and combinatorics*, volume 2213 of *Lecture notes in mathematics*. Springer, Cham, pp 317–323
- Sarnak P. Three lectures on the Möbius function, randomness and dynamics. Webpage. <http://publications.ias.edu/sarnak/>
- Tao T (2016) The logarithmically averaged Chowla and Elliott conjectures for two-point correlations. *Forum Math Pi* 4:e8, 36
- Tao T (2017) Equivalence of the logarithmically averaged Chowla and Sarnak conjectures. In: *Number theory – Diophantine problems, uniform distribution and applications*. Springer, Cham, pp 391–421
- Tao T (n.d.) The logarithmically averaged and non-logarithmically averaged Chowla conjectures. Webpage, <https://terrytao.wordpress.com/2017/10/20/the-logarithmically-averaged-and-non-logarithmically-averaged-chowlaconjectures/>
- Tao T, Teräväinen J (2018) Odd order cases of the logarithmically averaged Chowla conjecture. *J Théor Nombres Bordeaux* 30(3):997–1015
- Tao T, Teräväinen J (2019a) The structure of correlations of multiplicative functions at almost all scales, with applications to the Chowla and Elliott conjectures. *Algebra Number Theory* 13(9):2103–2150
- Tao T, Teräväinen J (2019b) The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures. *Duke Math J* 168(11):1977–2027
- Weiss B (1971) Normal sequences as collectives. In: *Proceedings of the symposium on topological dynamics and ergodic theory*. University of Kentucky

Further Reading

- Baake M, Huck C (2015) Ergodic properties of visible lattice points. *Proceedings of the Steklov Institute of Mathematics* 288(1):165–188
- Bergelson V, Ruzsa I (2002) Squarefree numbers, IP sets and ergodic theory. In *Paul Erdős and his mathematics, I* (Budapest, 1999), volume 11 of *Bolyai Soc. Math. Stud.*, pages 147–160. *János Bolyai Math. Soc.*, Budapest, 2002
- Bourgain J (2013) Möbius-Walsh correlation bounds and an estimate of Mauduit and Rivat. *J Anal Math* 119:147–163
- de la Bretèche R, Tenenbaum G (2019) A remark on Sarnak's conjecture. *Q J Math* 70(1):371–378
- Cellarosi F, Vinogradov I (2013) Ergodic properties of k -free integers in number fields. *J Mod Dyn* 7(3):461–488
- Dartyge C, Tenenbaum G (2005) Sommes des chiffres de multiples d'entiers. *Ann Inst Fourier (Grenoble)* 55(7):2423–2474
- Downarowicz T, Kasjan S (2015) Odometers and Toeplitz systems revisited in the context of Sarnak's conjecture. *Studia Math* 229(1):45–72
- Drmota M (2014) Subsequences of automatic sequences and uniform distribution. In *Uniform distribution and quasi-Monte Carlo methods*, volume 15 of *Radon Ser. Comput. Appl. Math.* De Gruyter, Berlin, pp 87–104
- Drmota M, Müllner C, Spiegelhofer L (2018) Möbius orthogonality for the Zeckendorf sum-of-digits function. *Proc Amer Math Soc* 146(9):3679–3691
- Dymek A (2017) Automorphisms of Toeplitz B -free systems. *Bull Pol Acad Sci Math* 65(2):139–152
- Dymek A, Kasjan S, Kułaga-Przymus J, Lemńczyk M (2018) B -free sets and dynamics. *Trans Amer Math Soc* 370(8):5425–5489
- Eisner T (2015) A polynomial version of Sarnak's conjecture. *C. R. Math Acad Sci Paris* 353(7):569–572
- El Abdalaoui EH. Oscillating sequences, Gowers norms and Sarnak's conjecture. Preprint, <https://arxiv.org/abs/1704.07243>
- El Abdalaoui EH, Askri G, Marzougui H. (2018) Möbius disjointness conjecture for local dendrite maps. *Nonlinearity* 32(1):285–300
- El Abdalaoui EH, Lemńczyk M, de la Rue T (2015) A dynamical point of view on the set of B -free integers. *Int Math Res Not* 2015(16):7258–7286
- Fan A-H, Jiang Y (2018) Oscillating sequences, MMA and MMLS flows and Sarnak's conjecture. *Ergodic Theory Dynam. Systems* 38(5):1709–1744
- Ferenczi S, Kułaga-Przymus J, Lemńczyk M, Mauduit C (2016) Substitutions and Möbius disjointness. In *Ergodic Theory, Dynamical Systems, and the Continuing Influence of John C. Oxtoby*, volume 678 of *Contemp. Math.* Amer. Math. Soc., Providence, RI, pp 151–173
- Flaminio L, Fraczek K, Kułaga-Przymus J, Lemńczyk M (2019) Approximate orthogonality of powers for ergodic affine unipotent diffeomorphisms on nilmanifolds. *Studia Math* 244(1):43–97
- Forni G, Kanigowski A (2020) Multiple mixing and disjointness for time changes of bounded-type Heisenberg nilflows. *Journal de l'École polytechnique – Mathématiques* 7:63–91
- Frantzikinakis N (2018) An averaged Chowla and Elliott conjecture along independent polynomials. *Int Math Res Not* 2018(12):3721–3743
- Green B (2012) On (not) computing the Möbius function using bounded depth circuits. *Combin Probab Comput* 21(6):942–951
- He X, Wang Z. Möbius disjointness for nilsequences along short intervals. Preprint, <https://arxiv.org/abs/1905.02864>
- Huang W, Lian Z, Shao S, Ye X (2017) Sequences from zero entropy noncommutative toral automorphisms and Sarnak Conjecture. *J Differential Equations* 263(1):779–810
- Huck C (2018) On the logarithmic probability that a random integral ideal is A -free. In *Ferenczi S, Kułaga-Przymus J, Lemńczyk M (eds) Ergodic theory and dynamical systems in their interactions with arithmetics and combinatorics*, volume 2213 of *Lecture notes in mathematics*. Springer, Cham, pp 249–258
- Kasjan S, Keller G, Lemńczyk M (2019) Dynamics of B -free sets: a view through the window. *Int Math Res Not* 2019(9):2690–2734
- Keller G. Generalized heredity in B -free systems. Preprint, <https://arxiv.org/abs/1704.04079>
- Konieczny J (2020) Möbius orthogonality for q -semi-multiplicative sequences. *Monatshefte Math* 192(4):853–882
- Konieczny J, Kupsa M, Kwietniak D (2018) Arcwise connectedness of the set of ergodic measures of hereditary shifts. *Proc Amer Math Soc* 146(8):3425–3438. <https://arxiv.org/abs/1610.00672>
- Kułaga-Przymus J, Lemńczyk M (2019) Möbius disjointness along ergodic sequences for uniquely ergodic actions. *Ergodic Theory Dynam Systems* 39(10):2793–2826
- Kułaga-Przymus J, Lemńczyk M, Weiss B (2015) On invariant measures for B -free systems. *Proc Lond Math Soc* (3), 110(6):1435–1474
- Kułaga-Przymus J, Lemńczyk M, Weiss B (2016) Hereditary subshifts whose simplex of invariant measures is Poulsen. In *Ergodic theory, dynamical systems, and the continuing influence of John C. Oxtoby*, volume 678 of *Contemp. Math.* Amer. Math. Soc. Providence, RI, pp 245–253
- Liu J, Sarnak P (2015) The Möbius function and distal flows. *Duke Math J* 164(7):1353–1399
- Matomäki K, Radziwiłł M, Tao T (2016) Sign patterns of the Liouville and Möbius functions. *Forum of Mathematics, Sigma*, 4

- Mauduit C, Rivat J (2015) Prime numbers along Rudin–Shapiro sequences. *J Eur Math Soc (JEMS)* 17 (10):2595–2642
- McNamara R. Sarnak's conjecture for sequences of almost quadratic word growth. Preprint, <https://arxiv.org/abs/1901.06460>
- Mentzen MK (2017) Automorphisms of subshifts defined by B-free sets of integers. *Colloq Math* 147(1):87–94
- Mirsky L (1947) Note on an asymptotic formula connected with r-free integers. *Quart J Math Oxford Ser* 18:178–182
- Mirsky L (1949) Arithmetical pattern problems relating to divisibility by rth powers. *Proc London Math Soc* (2) 50:497–508
- Peckner R (2018) Möbius disjointness for homogeneous dynamics. *Duke Math J* 167(14):2745–2792
- Pleasants PAB, Huck C (2013) Entropy and diffraction of the k-free points in n-dimensional lattices. *Discrete Comput Geom* 50(1):39–68
- Ryzhikov VV (2013) Bounded ergodic constructions, disjointness, and weak limits of powers. *Trans Moscow Math Soc* 165–171
- Sarnak P (2012) Möbius randomness and dynamics. *Not S Afr Math Soc* 43(2):89–97
- Sawin W. Dynamical models for Liouville and obstructions to further progress on sign patterns. Preprint, <https://arxiv.org/abs/1809.03280>
- Soundararajan K (2017) The Liouville function in short intervals. *Astérisque*, 2015/2016(390):Exp. No. 1119, 453–479. *Séminaire Bourbaki*. Vol. 2015/2016. *Exposés* 1104–1119
- Sun W. Sarnak's conjecture for nilsequences on arbitrary number fields and applications. Preprint, <https://arxiv.org/abs/1902.09712>
- Tao T. Furstenberg limits of the Liouville function. Webpage, <https://terrytao.wordpress.com/2017/03/05/furstenberg-limits-of-the-liouville-function/>
- Teräväinen J (2018) On binary correlations of multiplicative functions. *Forum Math Sigma* 6:e10, 41
- Veech WA (2017) Möbius orthogonality for generalized Morse-Kakutani flows. *Am J Math* 139(5):1157–1203
- Wang Z (2017) Möbius disjointness for analytic skew products. *Invent Math* 209(1):175–196



Smooth Ergodic Theory

Amie Wilkinson

Northwestern University, Evanston, IL, USA

Article Outline

Glossary

Definition of the Subject

Introduction

The Volume Class

The Fundamental Questions

Lebesgue Measure and Local Properties of
Volume

Ergodicity of the Basic Examples

Hyperbolic Systems

Beyond Uniform Hyperbolicity

The Presence of Critical Points and Other
Singularities

Future Directions

Bibliography

Glossary

Conservative, dissipative Conservative dynamical systems (on a compact phase space) are those that preserve a finite measure equivalent to volume. Hamiltonian dynamical systems are important examples of conservative systems. Systems that are not conservative are called dissipative. Finding physically meaningful invariant measures for dissipative maps is a central object of study in smooth ergodic theory.

Distortion estimate A key technique in smooth ergodic theory, a distortion estimate for a smooth map f gives a bound on the variation of the jacobian of f^n in a given region, for n arbitrarily large. The jacobian of a smooth map at a point x is the absolute value of the determinant of derivative at x , measured in a fixed Riemannian metric. The jacobian

measures the distortion of volume under f in that metric.

Hopf argument A technique developed by Eberhard Hopf for proving that a conservative diffeomorphism or flow is ergodic. The argument relies on the Ergodic Theorem for invertible transformations, the density of continuous functions among integrable functions, and the existence of stable and unstable foliations for the system. The argument has been used, with various modifications, to establish ergodicity for hyperbolic, partially hyperbolic and non-uniformly hyperbolic systems.

Hyperbolic A compact invariant set $\Lambda \subset M$ for a diffeomorphism $f: M \rightarrow M$ is hyperbolic if, at every point in Λ , the tangent space splits into two subspaces, one that is uniformly contracted by the derivative of f , and another that is uniformly expanded. Expanding maps and Anosov diffeomorphisms are examples of globally hyperbolic maps. Hyperbolic diffeomorphisms and flows are the archetypical smooth systems displaying chaotic behavior, and their dynamical properties are well-understood. Nonuniform hyperbolicity and partial hyperbolicity are two generalizations of hyperbolicity that encompass a broader class of systems and display many of the chaotic features of hyperbolic systems.

Sinai–Ruelle–Bowen (SRB) measure The concept of SRB measure is a rigorous formulation of what it means for an invariant measure to be “physically meaningful”. An SRB measure attracts a large set of orbits into its support, and its statistical features are reflected in the behavior of these attracted orbits.

Definition of the Subject

Smooth ergodic theory is the study of the statistical and geometric properties of measures invariant under a smooth transformation or flow. The study of smooth ergodic theory is as old as the study of

abstract ergodic theory, having its origins in Boltzmann's Ergodic Hypothesis in the late nineteenth Century. As a response to Boltzmann's hypothesis, which was formulated in the context of Hamiltonian Mechanics, Birkhoff and von Neumann defined ergodicity in the 1930s and proved their foundational ergodic theorems. The study of ergodic properties of smooth systems saw an advance in the work of Hadamard and E. Hopf in the 1930s their study of geodesic flows for negatively curved surfaces. Beginning in the 1950s, Kolmogorov, Arnold and Moser developed a perturbative theory producing obstructions to ergodicity in Hamiltonian systems, known as Kolmogorov–Arnold–Moser (KAM) Theory. Beginning in the 1960s with the work of Anosov and Sinai on hyperbolic systems, the study of smooth ergodic theory has seen intense activity. This activity continues today, as the ergodic properties of systems displaying weak forms of hyperbolicity are further understood, and KAM theory is applied in increasingly broader contexts.

Introduction

This entry focuses on the basic arguments and principles in smooth ergodic theory, illustrating with simple and straightforward examples. The classic texts (Arnol'd and Avez 1986; Mañé 1987) are a good supplement.

The discussion here sidesteps the topic of Kolmogorov–Arnold–Moser (KAM) Theory, which has played an important role in the development of smooth ergodic theory. For reasons of space, detailed discussion of several active areas in smooth ergodic theory is omitted, including: higher mixing properties (Kolmogorov, Bernoulli, etc.), finer statistical properties (fast decay of correlations, Central Limit Theorem, large deviations), smooth thermodynamic formalism (transfer operators, pressure, dynamical zeta functions, etc.), the smooth ergodic theory of random dynamical systems, as well as any mention of infinite invariant measures. The text (Baladi 2000) covers many of these topics, and the texts (Kifer 1986, 1988; Liu and Qian 1995) treat random smooth ergodic theory in depth. An excellent

discussion of many of the recent developments in the field of smooth ergodic theory is (Bonatti et al. 2005).

This entry assumes knowledge of the basic concepts in ergodic theory and of basic differential topology. The texts (Cornfeld et al. 1982) and (Hirsch 1979) contain the necessary background.

The Volume Class

For simplicity, assume that M is a compact, boundaryless C^∞ Riemannian manifold, and that $f: M \rightarrow M$ is an orientation-preserving, C^1 map satisfying $m(D_x f) > 0$, for all $x \in M$, where

$$m(D_x f) = \inf_{v \in T_x M, \|v\|=1} \|D_x f(v)\|.$$

If f is a diffeomorphism, then this assumption is automatically satisfied, since in that case $m(D_x f) = \|D_{f(x)} f^{-1}\|^{-1} > 0$. For non-invertible maps, this assumption is essential in much of the following discussion. The Inverse Function Theorem implies that any map f satisfying these hypotheses is a covering map of positive degree $d \geq 1$.

These assumptions will avoid the issues of infinite measures and the behavior of f near critical points and singularities of the derivative. For most results discussed in this entry, this assumption is not too restrictive. The existence of critical points and other singularities is, however, a complication that cannot be avoided in many important applications. The ergodic-theoretic analysis of such examples can be considerably more involved, but contains many of the elements discussed in this entry. The discussion in section “Beyond Uniform Hyperbolicity” indicates how some of these additional technicalities arise and can be overcome. For simplicity, the discussion here is confined almost exclusively to discrete time evolution. Many, though not all, of the results mentioned here carryover to flows and semi flows using, for example, a cross-section construction (see Chap. 1 in (Mañé 1987)).

Every smooth map $f: M \rightarrow M$ satisfying these hypotheses preserves a natural measure *class*, the

measure class of a finite, smooth Riemannian volume on M . Fix such a volume ν on M . Then there exists a continuous, positive *Jacobian* function $x \mapsto \text{jac}_x f$ on M , with the property that for every sufficiently small ball $B \subset M$, and every measurable set $A \subset B$ one has:

$$\nu(f(A)) = \int_B \text{jac}_x f \, d\nu(x) .$$

The Jacobian of f at x is none other than the absolute value of the determinant of the derivative $D_x f$ (measured in the given Riemannian metric). To see that the measure class of ν is preserved by f , observe that the Radon–Nikodym derivative $\frac{df_* \nu}{d\nu}(x)$ at x is equal to $\sum_{y \in f^{-1}(x)} (\text{jac}_y f)^{-1} > 0$. Hence $f_* \nu$ is equivalent to ν , and f preserves the measure class of ν .

In many contexts, the map f has a natural *invariant* measure in the measure class of volume. In this case, f is said to be *conservative*. One setting in which a natural invariant smooth measure appears is Hamiltonian dynamics. Any solution to Hamilton's equations preserves a smooth volume called the *Liouville measure*. Furthermore, along the invariant, constant energy hyper manifolds of a Hamiltonian flow, the Liouville measure decomposes smoothly into invariant measures, each of which is equivalent to the induced Riemannian volume. In this way, many systems of physical or geometric origin, such as billiards, geodesic flows, hard sphere gases, and evolution of the n -body problem give rise to smooth conservative dynamical systems. See Dynamics of Hamiltonian Systems.

Note that even though f preserves a smooth measure class, it might not preserve any measure in that measure class. Consider, for example, a diffeomorphism $f: S^1 \rightarrow S^1$ of the circle with exactly two fixed points, p and q , $f'(p) > 1 > f'(q) > 0$. Let μ be an f -invariant probability measure. Let I be a neighborhood of p . Then $\bigcap_{n=1}^{\infty} f^{-n}(I) = \{p\}$, but on the other hand, $\mu(f^{-n}(I)) = \mu(I) > 0$, for all n . This implies that $\mu(\{p\}) > 0$, and so μ does not lie in the measure class of volume. This is an example of a

dissipative map. A map f is called *dissipative* if every f -invariant measure with full support has a singular part with respect to volume. As was just seen, if a diffeomorphism f has a periodic sink, then f is dissipative; more generally, if a diffeomorphism f has a periodic point p of period k such that $\text{jac}_p f^k \neq 1$, then f is dissipative.

The Fundamental Questions

For a given smooth map $f: M \rightarrow M$, there are the following fundamental questions.

1. Is f conservative? That is, does there exist an invariant measure in the class of volume? If so, is it unique?
2. When f is conservative, what are its statistical properties? Is it ergodic, mixing, a K-system, Bernoulli, etc.? Does it obey a Central Limit Theorem, fast decay of correlations, large deviations estimates, etc.?
3. If f is dissipative, does there exist an invariant measure, not in the class of volume, but (in some sense) natural with respect to volume? What are the statistical properties of such a measure, if it exists?

There are several plausible ways to “answer” these questions. One might fix a given map f of interest and ask these questions for that specific f . What tends to happen in the analysis of a single map f is that either:

- the question can be answered using “soft” methods, and so the answer applies not only to f but to perturbations of f , or even to *generic* or *typical* f inside a class of maps; or,
- the proof requires “hard” analysis or precise asymptotic information and cannot possibly be answered for a specific f , but can be answered for a large set of f_i in a typical (or given) parametrized family $\{f_t\}_{t \in (-1, 1)}$ of smooth maps containing $f = f_0$.

Both types of results appear in the discussion that follows.

Lebesgue Measure and Local Properties of Volume

Locally, any measure in the measure class of volume is, after a smooth change of coordinates, equivalent to Lebesgue measure in $\mathbb{R}^n$. In fact, more is true: Moser's Theorem implies that locally any Riemannian volume is, after a smooth change of coordinates, *equal* to Lebesgue measure in $\mathbb{R}^n$. Hence to study many of the local properties of volume, it suffices to study the same properties for Lebesgue measure.

One of the basic properties of Lebesgue measure is that every set of positive Lebesgue measure can be approximated arbitrarily well in measure from the outside by an open set, and from the inside by a compact set. A consequence of this property, of fundamental importance in smooth ergodic theory, is the following statement.

Fundamental Principle # 1: Two disjoint, positive Lebesgue measure sets cannot mix together uniformly at all scales.

As an illustration of this principle, consider the following elementary exercise in measure theory. First, some notation. If ν is a measure and A and B are ν -measurable sets with $\nu(B) > 0$, the *density* of A in B is defined by:

$$\nu(A : B) = \frac{\nu(A \cap B)}{\nu(B)} .$$

Proposition 1 *Let $\mathcal{P}_1, \mathcal{P}_2, \dots$ be sequence of (mod 0) finite partitions of the circle S^1 into open intervals, with the properties: (a) any element of $\mathcal{P}_n$ is a (mod 0) union of elements of $\mathcal{P}_{n+1}$, and (b) the maximum diameter of elements of $\mathcal{P}_n$ tends to 0 as $n \rightarrow \infty$.*

Let A be any set of positive Lebesgue measure in S^1 . Then there exists a sequence of intervals $I_1, I_2, \dots$, with $I_n \in \mathcal{P}_n$ such that:

$$\lim_{n \rightarrow \infty} \lambda(A : I_n) = 1 .$$

Proof Assume that Lebesgue measure has been normalized so that $\lambda(S^1) = 1$. Fix a (mod 0) cover

of $S^1 \setminus A$ by pairwise disjoint elements $\{J_i\}$ of the union $\bigcup_{n=1}^{\infty} \mathcal{P}_n$ with the properties:

$$\lambda(J_1) \geq \lambda(J_2) \geq \dots, \text{ and}$$

$$\lambda\left(\bigcup_{i=1}^{\infty} J_i\right) = \sum_{i=1}^{\infty} \lambda(J_i) < 1 .$$

For $n \in \mathbb{N}$, let U_n be the union of all the intervals J_i that are contained in $\mathcal{P}_n$, and let $V_n = \bigcup_{i=1}^n U_i$. This defines an increasing sequence of natural numbers $i_1 = 1 < i_2 < i_3 < \dots$ such that $U_n = \bigcup_{i=i_n+1}^{i_{n+1}-1} J_i$ and $V_n = \bigcup_{i=1}^{i_{n+1}-1} J_i$.

For each n , the interval I_n will be chosen to be an element $\mathcal{P}_n$, disjoint from V_n , in which the density of $\bigcup_{i=n+1}^{\infty} U_i$ is very small (approaching 0 as $n \rightarrow \infty$). Since $(S^1 \setminus A) \cap I_n$ is contained in $\bigcup_{i=n+1}^{\infty} U_i$, this choice of I_n will ensure that the density of A in I_n is large (approaching 1 as $n \rightarrow \infty$).

To make this choice of I_n , note first that the density of $\bigcup_{i=n+1}^{\infty} U_i$ inside of $S^1 \setminus V_n$ is:

$$\frac{\lambda\left(\bigcup_{i=n+1}^{\infty} U_i\right)}{\lambda(S^1 \setminus V_n)} = \frac{\sum_{i=i_{n+1}}^{\infty} \lambda(J_i)}{1 - \sum_{i=1}^{i_{n+1}-1} \lambda(J_i)} = a_n .$$

Note that, since $\sum_{i=1}^{\infty} \lambda(J_i) < 1$, one has $a_n \rightarrow 0$ as $n \rightarrow \infty$. Since the density of $\bigcup_{i=n+1}^{\infty} U_i$ inside of $S^1 \setminus V_n$ is at most a_n , there is an interval I_n in $\mathcal{P}_n$, disjoint from V_n , such that the density of $\bigcup_{i=n+1}^{\infty} U_i$ inside of I_n is at most a_n . Then

$$\lim_{n \rightarrow \infty} \lambda(A : I_n) \geq \lim_{n \rightarrow \infty} 1 - a_n = 1 .$$

□

In smooth ergodic theory, it is often useful to use a variation on Proposition 1 (generally, in higher dimensions) in which the partitions $\mathcal{P}_n$ are nested, dynamically-defined partitions. A simple application of this method can be used to prove that the doubling map on the circle is ergodic with respect to Lebesgue measure, which is done in section “[Lebesgue Measure and Local Properties of Volume](#)”.

Notice that this proposition does not claim that the intervals I_n are nested. If one imposes stronger

conditions on the partitions $\mathcal{P}_n$, then one can draw stronger conclusions.

A very useful theorem in this respect is the Lebesgue Density Theorem. A point $x \in M$ is a *Lebesgue density point* of a measurable set $X \subseteq M$ if

$$\lim_{r \rightarrow 0} m(X : B_r(x)) = 1 ,$$

where $B_r(x)$ is the Riemannian ball of radius r centered at x . Notice that the notion of Lebesgue density point depends only on the smooth structure of M , because any two Riemannian metrics have the same Lebesgue density points. The Lebesgue Density Theorem states that if A is a measurable set and $\hat{A}$ is the set of Lebesgue density points of A , then $m(A \Delta \hat{A}) = 0$.

Ergodicity of the Basic Examples

This section contains proofs of the ergodicity of two basic examples of conservative smooth maps: irrational rotations on the circle and the doubling map on the circle. See ► [Ergodic Theory: Basic Examples and Constructions](#) for a more detailed description of these maps. These proofs serve as an elementary illustration of some of the fundamental techniques and principles in smooth ergodic theory.

Rotations on the circle. Denote by S^1 the circle $\mathbb{R}/\mathbb{Z}$, which is an additive group, and by λ normalized Lebesgue-Haar measure on S^1 . Fix a real number $\alpha \in \mathbb{R}$. The rotation $R_\alpha : S^1 \rightarrow S^1$ is the translation defined by $R_\alpha(x) = x + \alpha$. Since translations preserve Lebesgue-Haar measure, the map R_α is conservative. Note that R_α is a diffeomorphism and an isometry with respect to the canonical flat metric (length) on S^1 .

Proposition 2 If $\alpha \notin \mathbb{Q}$, then the rotation $R_\alpha : S^1 \rightarrow S^1$ is ergodic with respect to Lebesgue measure.

Proof Let A be an R_α -invariant set in S^1 , and suppose that $0 < \lambda(A) < 1$. Denote by A^c the complement of A in S^1 . Fix $\varepsilon > 0$. Proposition 1

implies that there exists an interval $I \subset S^1$ such that the density of A in I is large: $\lambda(A : I) > 1 - \varepsilon$. Similarly, one may choose an interval J such that $\lambda(A^c : J) > 1 - \varepsilon$. Without loss of generality, one may choose I and J to have the same length. Since α is irrational, R_α has a dense orbit, which meets the interval I . Since R_α is an isometry, this implies that there is an integer n such that $\lambda(R_\alpha^n(I) \Delta J) < \varepsilon \lambda(I)$. Since $\lambda(I) = \lambda(J)$, this readily implies that $|\lambda(A : R_\alpha^n(I)) - \lambda(A : J)| < \varepsilon$. Also, since A is invariant, and R_α is invertible and preserves measure, one has:

$$\begin{aligned} \lambda(A : R_\alpha^n(I)) &= \lambda(R_\alpha^n(A) : R_\alpha^n(I)) = \lambda(A : I) \\ &> 1 - \varepsilon . \end{aligned}$$

But for ε sufficiently small, this contradicts the facts that $\lambda(A : J) = 1 - \lambda(A^c : J) < \varepsilon$ and $|\lambda(A : R_\alpha^n(I)) - \lambda(A : J)| < \varepsilon$. $\square$

Note that this is not a proof of the strongest possible statement about R_α (namely, minimality and unique ergodicity). The point here is to show how “soft” arguments are often sufficient to establish ergodicity; this proof uses no more about R_α than the fact that it is a transitive isometry. Hence the same argument shows:

Theorem 1 Let $f : M \rightarrow M$ be a transitive isometry of a Riemannian manifold M . Then f is ergodic with respect to Riemannian volume.

One can isolate from this proof a useful principle:

Fundamental Principle # 2: Isometries preserve Lebesgue density at all scales, for arbitrarily many iterates.

This principle implies, for example, that a smooth action by a compact Lie group on M is ergodic along typical (nonsingular) orbits. This principle is also useful in studying area-preserving flows on surfaces and, in a refined form, unipotent flows on homogeneous spaces. In the case of surface flows, ergodicity questions can be reduced to a study of interval exchange transformations. See the entry ► [Ergodic Theory: Basic Examples and Constructions](#) for a detailed discussion of interval exchange transformations and flows on

surfaces. ► [Introduction to Ergodic Theory](#) contains detailed information on unipotent flows.

Doubling map on the circle. Let $T_2 : S^1 \rightarrow S^1$ be the doubling map defined by $T_2(x) = 2x$. Then T_2 is a degree-2 covering map and endomorphism of S^1 with constant jacobian $\text{jac}_x T_2 \equiv 2$. Since $\frac{d(T_2)_* \lambda}{d\lambda} = \frac{1}{2} + \frac{1}{2} = 1$, T_2 preserves Lebesgue-Haar measure. The doubling map is the simplest example of a hyperbolic dynamical system, a topic treated in depth in the next section.

As with the previous example, the focus here is on the property of ergodicity. It is again possible to prove much stronger results about T_2 , such as Bernoulli city, by other methods. Instead, here is a soft proof of ergodicity that will generalize readily to other contexts.

Proposition 3 *The doubling map $T_2 : S^1 \rightarrow S^1$ is ergodic with respect to Lebesgue measure.*

Proof Let A be a T_2 -invariant set in S^1 with $\lambda(A) > 0$. Let $p \in S^1$ be the fixed point of T_2 , so that $T_2(p) = p$. For each $n \in \mathbb{N}$, the preimages of p under T_2^{-n} define a (mod 0) partition $\mathcal{P}_n$ into 2^n open intervals of length 2^{-n} ; the elements of $\mathcal{P}_n$ are the connected components of $S^1 \setminus T_2^{-n}(\{p\})$. Note that the sequence of partitions $\mathcal{P}_1, \mathcal{P}_2, \dots$ is nested, in the sense of Proposition 1. Restricted to any interval $J \in \mathcal{P}_n$, the map T_2^n is a diffeomorphism onto $S^1 \setminus \{p\}$ with constant jacobian $\text{jac}_x(T_2^n) = (T_2^n)'(x) = 2^n$.

Since A is invariant, it follows that $T_2^{-n}(A) = A$. Fix $\varepsilon > 0$. Proposition 1 implies that there exists an $n \in \mathbb{N}$ and an interval $J \in \mathcal{P}_n$ such that $\lambda(A : J) > 1 - \varepsilon$. Note that $T_2^n(A \cap J) \subset A$. But then

$$\begin{aligned} \lambda(A) &\geq \lambda(T_2^n(A \cap J)) \\ &= \int_{A \cap J} \text{jac}_x(T_2^n) d\lambda(x) \\ &= 2^n \lambda(A \cap J) \\ &= 2^n \lambda(A : J) \lambda(J) \\ &> 2^n (1 - \varepsilon) \lambda(J) = 1 - \varepsilon. \end{aligned}$$

Since ε was arbitrary, one obtains that $\lambda(A) = 1$. $\square$

In this proof, the facts that the intervals in $\mathcal{P}_n$ have constant length 2^{-n} and that the jacobian of T_2^n restricted to such an interval is constant and equal to 2^n are not essential. The key fact really used in this proof is the assertion that the ratio:

$$\frac{\lambda(T_2^n(A \cap J) : T_2^n(J))}{\lambda(A : J)}$$

is bounded, *independently of n* . In this case, the ratio is 1 for all n because T_2 has constant jacobian.

It is tempting to try to extend this proof to other expanding maps on the circle, for example, a C^1 , λ -preserving map $f : S^1 \rightarrow S^1$ with $d_{C^1}(f, T_2, \cdot)$ small. Many of the aspects of this proof carry through *mutatimutandis* for such an f , save for one. A C^1 -small perturbation of T_2 will in general no longer have constant jacobian, and the *variation* of the jacobian of f^n on a small interval can be (and often is) unbounded. The reason for this unboundedness is a lack of control of the modulus of continuity of f' . Hence this argument can fail for C^1 perturbations of T_2 . On the other hand, the argument still works for C^2 perturbations of T_2 , even when the jacobian is not constant.

The principle behind this fact can be loosely summarized:

Fundamental Principle # 3: On controlled scales, iterates of C^2 expanding maps distort Lebesgue density in a controlled way.

This principle requires further explanation and justification, which will come in the following section. The C^2 hypothesis in this principle accounts for the fact that almost all results in smooth ergodic theory assume a C^2 hypothesis (or something slightly weaker).

Hyperbolic Systems

One of the most developed areas of smooth ergodic theory is in the study of hyperbolic maps and attractors. This section defines hyperbolic maps and attractors, provides examples, and investigates their ergodic properties. See (Robinson

1995; Katok and Hasselblatt 1995) and Hyperbolic Dynamical Systems for a thorough discussion of the topological and smooth properties of hyperbolic systems.

A *hyperbolic structure* on a compact f -invariant set $\Lambda \subset M$ is given by a Df -invariant splitting $T_\Lambda M = E^u \oplus E^s$ of the tangent bundle over Λ and constants $C, \mu > 1$ such that, for every $x \in \Lambda$ and $n \in \mathbb{N}$:

$$\begin{aligned} v \in E^u(x) &\Rightarrow \|D_x f^n(v)\| \geq C^{-1} \mu^n \|v\|, \\ \text{and } v \in E^s(x) &\Rightarrow \|D_x f^n(v)\| \leq C \mu^{-n} \|v\|. \end{aligned}$$

A hyperbolic attractor for a map $f: M \rightarrow M$ is given by an open set $U \subset M$ such that: $f(U) \subset \overline{U}$, and such that the set $\Lambda = \bigcap_{n \geq 0} f^n(U)$ carries a hyperbolic structure. The set Λ is called the *attractor*, and U is an *attracting region*. A map $f: M \rightarrow M$ is *hyperbolic* if M decomposes (mod 0) into a finite union of attracting regions for hyperbolic attractors. Typically one assumes as well that the restriction of f to each attractor Λ_i is topologically transitive.

Every point p in a hyperbolic set Λ has smooth *stable manifold* $\mathcal{W}^s(p)$ and *unstable manifold* $\mathcal{W}^u(p)$, tangent, respectively, to the subspaces $E^s(p)$ and $E^u(p)$. The set $\mathcal{W}^s(p)$ is precisely the set of $q \in M$ such that $d(f^n(p), f^n(q))$ tends to 0 as $n \rightarrow \infty$, and it follows that $f(\mathcal{W}^s(p)) = \mathcal{W}^s(f(p))$. When f is a diffeomorphism, the unstable manifold $\mathcal{W}^u(p)$ is uniquely defined and is the stable manifold of f^{-1} . When f is not invertible, local unstable manifolds exist, but generally are not unique. If Λ is a transitive hyperbolic attractor, then every unstable manifold of every point $p \in \Lambda$ is dense in Λ .

Examples of Hyperbolic Maps and Attractors

Expanding Maps

The previous section mentioned briefly the C^r perturbations of the doubling map T_2 . Such perturbations (as well as T_2 itself) are examples of *expanding maps*. A map $f: M \rightarrow M$ is *expanding* if there exist constants $\mu > 1$ and $C > 0$ such that, for every $x \in M$, and every nonzero vector $v \in T_x M$:

$$\|D_x f^n(v)\| \geq C \mu^n \|v\|,$$

with respect to some (any) Riemannian metric on M . An expanding map is clearly hyperbolic, with $U = M$, E^s the trivial bundle, and $E^u = TM$. Any disk in M is a local unstable manifold for f .

Anosov Diffeomorphisms

A diffeomorphism $f: M \rightarrow M$ is called *Anosov* if the tangent bundle splits as a direct sum $TM = E^u \oplus E^s$ of two Df -invariant subbundles, such that E^u is uniformly expanded and E^s is uniformly contracted by Df . Similarly, a flow $\varphi_t: M \rightarrow M$ is called Anosov if the tangent bundle splits as a direct sum $TM = E^u \oplus E^0 \oplus E^s$ of three $D\varphi_t$ -invariant subbundles, such that E^0 is generated by $\dot{\varphi}$, E^u is uniformly expanded and E^s is uniformly contracted by $D\varphi_t$. Like expanding maps, an Anosov diffeomorphism is an Anosov attractor with $\Lambda = U = M$.

A simple example of a conservative Anosov diffeomorphism is a hyperbolic linear automorphism of the torus. Any matrix $A \in SL(n, \mathbb{Z})$ induces an automorphism of $\mathbb{R}^n$ preserving the integer lattice $\mathbb{Z}^n$, and so descends to an automorphism $f_A: T^n \rightarrow T^n$ of the n -torus $T^n = \mathbb{R}^n/\mathbb{Z}^n$. Since the determinant of A is 1, the diffeomorphism f_A preserves Lebesgue-Haar measure on T^n . In the case where none of the eigenvalues of A have modulus 1, the resulting diffeomorphism f_A is Anosov. The stable bundle E^s at $x \in T^n$ is the parallel translate to x of the sum of the contracted generalized eigenspaces of A , and the unstable bundle E^u at x is the translated sum of expanded eigenspaces.

In general, the invariant subbundles E^u and E^s of an Anosov diffeomorphism are integrable and tangent to a transverse pair of foliations $\mathcal{W}^u$ and $\mathcal{W}^s$, respectively (see, e.g. (Hirsch et al. 1977). for a proof of this). The leaves of $\mathcal{W}^s$ are uniformly contracted by f , and the leaves of $\mathcal{W}^u$ are uniformly contracted by f^{-1} . The leaves of these foliations are as smooth as f , but the tangent bundles to the leaves do not vary smoothly in the manifold. The regularity properties of these foliations play an important role in the ergodic properties of Anosov diffeomorphisms.

The first Anosov flows to be studied extensively were the geodesic flows for manifolds of negative sectional curvatures. As these flows are Hamiltonian, they are conservative. Eberhard Hopf showed in the 1930s that such geodesic flows for surfaces are ergodic with respect to Liouville measure (Hopf 1939); it was not until the 1960s that ergodicity of all such flows was proved by Anosov (Anosov 1967). The next section describes, in the context of Anosov diffeomorphisms, Hopf's method and important refinements due to Anosov and Sinai.

DA Attractors

A simple way to produce a non-Anosov hyperbolic attractor on the torus is to start with an Anosov diffeomorphism, such as a linear hyperbolic automorphism, and deform it in a neighborhood of a fixed point, turning a saddle fixed point into a source, while preserving the stable foliation. If this procedure is carried out carefully enough, the resulting diffeomorphism is a dissipative hyperbolic diffeomorphism, called a *derived from Anosov (DA)* attractor. Other examples of hyperbolic attractors are the Plykin attractor and the solenoid. See (Robinson 1995).

Distortion Estimates

Before describing the ergodic properties of hyperbolic systems, it is useful to pause for a brief discussion of distortion estimates. Distortion estimates are behind almost every result in smooth ergodic theory. In the hyperbolic setting, distortion estimates are applied to the action of f on unstable manifolds to show that the volume distortion of f along unstable manifolds can be controlled for arbitrarily many iterates.

The example mentioned at the end of the previous section illustrates the ideas in a distortion estimate. Suppose that $f : S^1 \rightarrow S^1$ is a C^2 expanding map, such as a C^2 small perturbation of T_2 . Then there exist constants $\mu > 1$ and $C > 0$ such that $(f^n)'(x) > C\mu^n$ for all x and n .

Let d be the degree of f . If I is a sufficiently small open interval in S^1 , then for each n , $f^{-n}(I)$ is a union of d disjoint intervals. Furthermore, each of these intervals has diameter at most $C^{-1}\mu^{-n}$ times the diameter of I . It is now possible to justify

the assertion in Fundamental Principle # 3 in this context.

Lemma *There exists a constant $K \geq 1$ such that, for all $n \in \mathbb{N}$, and for all $x, y \in f^{-n}(I)$, one has:*

$$K^{-1} \leq \frac{(f^n)'(x)}{(f^n)'(y)} \leq K.$$

Proof Since f is C^2 and f' is bounded away from 0, the function $\alpha(x) = \log(f'(x))$ is C^1 . In particular, α is Lipschitz continuous: there exists a constant $L > 0$ such that, for all $x, y \in S^1$, $|\alpha(x) - \alpha(y)| < Ld(x, y)$. For $n \geq 0$, let $\alpha_n(x) = \log((f^n)'(x))$. The Chain Rule implies that $\alpha_n(x) = \sum_{i=0}^{n-1} \alpha(f^i(x))$.

The expanding hypothesis on f implies that for all $x, y \in f^n(I)$ and for $i = 0, \dots, n$, one has $d(f^i(x), f^i(y)) \leq C^{-1}\mu^{i-n}d(f^n(x), f^n(y)) \leq C^{-1}\mu^{i-n}$. Hence

$$\begin{aligned} |\alpha_n(x) - \alpha_n(y)| &\leq \sum_{i=0}^{n-1} |\alpha(f^i(x)) - \alpha(f^i(y))| \\ &\leq L \sum_{i=0}^{n-1} d(f^i(x), f^i(y)) \\ &\leq L \sum_{i=0}^{n-1} C^{-1}\mu^{i-n} \\ &< LC^{-1}\mu^{-1}(1 - \mu^{-1})^{-1}. \end{aligned}$$

Setting $K = \exp(LC^{-1}\mu^{-1}(1 - \mu^{-1})^{-1})$, one now sees that $(f^n)'(x)/(f^n)'(y)$ lies in the interval $[K^{-1}, K]$, proving the claim. $\square$

In this distortion estimate, the function $\alpha : M \rightarrow \mathbb{R}$ is called a *cocycle*. The same argument applies to any Lipschitz continuous (or even Hölder continuous) cocycle.

Ergodicity of Expanding Maps

The ergodic properties of C^2 expanding maps are completely understood. In particular, every conservative expanding map is ergodic, and every expanding map is conservative. The proofs of these facts use Fundamental Principles # 1 and 3 in a fairly direct way.

Every C^2 conservative expanding map is ergodic with respect to volume. The proof is a straightforward adaptation of the proof of Proposition 3 (see, e.g. (Mañé 1987)). Here is a description of the proof for $M = S^1$. As remarked earlier, the proof of Proposition 3 adapts easily to a general expanding map $f: S^1 \rightarrow S^1$ once one shows that for every f -invariant set A , and every connected component J of $f^{-n}(S^1 \setminus \{p\})$, the quantity

$$\frac{\lambda(f^n(A \cap J) : f^n(J))}{\lambda(A : J)}$$

is bounded independently of n . This is a fairly direct consequence of the distortion estimate in Lemma 6.1 and is left as an exercise.

The same distortion estimates show that every C^2 expanding map is conservative, preserving a probability measure ν in the measure class of volume. Here is a sketch of the proof for the case $M = S^1$. To prove this, consider the push-forward $\lambda_n = f_*^n \lambda$. Then λ_n is equivalent to Lebesgue, and its Radon–Nikodym derivative $d\lambda_n \setminus d\lambda$ is the density function

$$\rho_n(x) = \sum_{y \in f^{-n}(x)} \frac{1}{j_{ac,y} f^n}.$$

Since $f_*^n \lambda$ is a probability measure, it follows that $\int_{S^1} \rho_n d\lambda = 1$. A simple argument using the distortion estimate above (and summing up over all d^n branches of f^n at x) shows that there exists a constant $c \geq 1$ such that for all $x, y \in S^1$,

$$c^{-1} \leq \frac{\rho_n(x)}{\rho_n(y)} \leq c.$$

Since the integral of ρ_n is 1, the functions ρ_n are uniformly bounded away from 0 and ∞ . It is easy to see that the measure $\nu_n = \frac{1}{n} \sum_{i=1}^n f_*^i \lambda$ has density $\frac{1}{n} \sum_{i=1}^n \rho_i$. Let ν be any subsequential weak* limit of ν_n ; then ν is absolutely continuous, with density ρ bounded away from 0 and ∞ . With a little more care, one can show that ρ is actually Lipschitz continuous.

As a passing comment, the ergodicity of ν and positivity of ρ imply that ν is the unique

f -invariant measure absolutely continuous with respect to λ . With more work, one can show that ν is exact. See (Mañé 1987) for details.

Ergodicity of Conservative Anosov Diffeomorphisms

Like conservative C^2 expanding maps, conservative C^2 Anosov diffeomorphisms are ergodic. This subsection outlines a proof of this fact. Unlike expanding maps, however, Anosov diffeomorphisms need not be conservative. The subsection following this one describe a type of invariant measure that is “natural” with respect to volume, called a Sinai–Ruelle–Bowen (or SRB) measure. The central result for hyperbolic systems states that every hyperbolic attractor carries an SRB measure.

The Hopf Argument

In the 1930s Hopf (Hopf 1939) proved that the geodesic flow for a compact, negatively-curved surface is ergodic. His method was to study the Birkhoff averages of continuous functions along leaves of the stable and unstable foliations of the flow. This type of argument has been used since then in increasingly general contexts, and has come to be known as the Hopf Argument.

The core of the Hopf Argument is very simple. To any $f: M \rightarrow M$ one can associate the *stable equivalence relation* $\sim_s$, where $x \sim_s y$ iff $\lim_{n \rightarrow \infty} d(f^n(x), f^n(y)) = 0$. Denote by $W^s(x)$ the stable equivalence class containing x . When f is invertible, one defines the *unstable equivalence relation* to be the stable equivalence relation for f^{-1} , and one denotes by $W^u(x)$ the unstable equivalence class containing x .

The first step in the Hopf Argument is to show that Birkhoff averages for continuous functions are constant along stable and unstable equivalence classes. Let $\phi: M \rightarrow \mathbb{R}$ be an integrable function, and let

$$\bar{\phi} = \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \phi \circ f^i. \quad (1)$$

Observe that if ϕ is continuous, then for every $x \in M$ and $x' \in W^s(x)$, $\lim_{n \rightarrow \infty} |\phi(f^n(x)) -$

$\phi(f^i(x')) \mid = 0$. It follows immediately that $\overline{\phi}_f(x) = \overline{\phi}_f(x')$. In particular, if the limit in (1) exists at x , then it exists and is constant on $W^s(x)$.

Fundamental Principle # 4: Birkhoff averages of continuous functions are constant along stable equivalence classes.

The next step of Hopf's argument confines itself to the situation where f is conservative and Anosov. In this case, f is invertible, the stable equivalence classes are precisely the leaves of the stable foliation $\mathcal{W}^s$, and the unstable equivalence classes are the leaves of the unstable foliation $\mathcal{W}^u$. Since f is conservative, the Ergodic Theorem implies that for every L^2 function ϕ , the function $\overline{\phi}_f$ is equal (mod 0) to the projection of ϕ onto the f -invariant functions in L^2 . Since this projection is continuous, and the continuous functions are dense in L^2 , to prove that f is ergodic, it suffices to show that the projection of any continuous function is trivial. That is, it suffices to show that for every continuous ϕ , the function $\overline{\phi}_f$ is constant (a.e.).

To this end, let $\phi : M \rightarrow \mathbb{R}$ be continuous. Since the f -invariant functions coincide with the f^{-1} -invariant functions, one obtains that $\overline{\phi}_f = \overline{\phi}_{f^{-1}}$ a.e. The previous argument shows $\overline{\phi}_f$ is constant along $\mathcal{W}^s$ -leaves and $\overline{\phi}_{f^{-1}}$ is constant along $\mathcal{W}^u$ -leaves. The desired conclusion is that $\overline{\phi}_f$ is a.e. constant. It suffices to show this in a local chart, since the manifold M is connected. In a local chart, after a smooth change of coordinates, one obtains a pair of transverse foliations $\mathcal{F}_1, \mathcal{F}_2$ of the cube $[-1, 1]^n$ by disks, and a measurable function $\psi : [-1, 1]^n \rightarrow \mathbb{R}$ that is constant along the leaves of $\mathcal{F}_1$ and constant along the leaves of $\mathcal{F}_2$.

When the foliations $\mathcal{F}_1$ and $\mathcal{F}_2$ are smooth (at least C^1), one can perform a further smooth change of coordinates so that $\mathcal{F}_1$ and $\mathcal{F}_2$ are transverse coordinate subspace foliations. In this case, Fubini's theorem implies that any measurable function that is constant along two transverse coordinate foliations is a.e. constant. This completes the proof in the case that the foliations $\mathcal{W}^s$ and $\mathcal{W}^u$ are smooth. In Hopf's original argument, the stable and unstable foliations were assumed to

be C^1 foliations (a hypotheses satisfied in the examples he considered, due to low-dimensionality. See also (Sinai 1961), where a pinching condition on the curvature, rather than low dimensionality, implies this C^1 condition on the foliations.)

Absolute Continuity

For a general Anosov diffeomorphism or flow, the stable and unstable foliations are not C^1 , and so the final step in Hopf's original argument does not apply. The fundamental advance of Anosov and Anosov-Sinai was to prove that the stable and unstable foliations of an Anosov diffeomorphism (conservative or not) satisfy a weaker condition than smoothness, called *absolute continuity*. For conservative systems, absolute continuity is enough to finish Hopf's argument, proving that every C^2 conservative Anosov diffeomorphism is ergodic (Anosov 1967; Anosov and Sinai 1967).

For a definition and careful discussion of absolute continuity of a foliation $\mathcal{F}$, see (Brin and Stuck 2002). Two consequences of the absolute continuity of $\mathcal{F}$ are:

1. (AC1) If $A \subset M$ is any measurable set, then

$$\lambda(A) = 0 \Leftrightarrow \lambda_{\mathcal{F}(x)}(A) = 0, \\ \text{for } \lambda - a.e. x \in M,$$

where $\lambda_{\mathcal{F}(x)}$ denotes the induced Riemannian volume on the leaf of $\mathcal{F}$ through x .

2. (AC2) If τ is any small, smooth disk transverse to a local leaf of $\mathcal{F}$, and $T \subset \tau$ is a 0-set in τ (with respect to the induced Riemannian volume on τ), then the union of the $\mathcal{F}$ leaves through points in T has Lebesgue measure 0 in M .

The proof that $\mathcal{W}^s$ and $\mathcal{W}^u$ are absolutely continuous has a similar flavor to the proof that an expanding map has a unique absolutely continuous invariant measure (although the cocycles involved are Hölder continuous, rather than Lipschitz), and the facts are intimately related.

With absolute continuity of the stable and unstable foliations in hand, one can now prove:

Theorem 2 (Anosov) *Let f be a C^2 , conservative Anosov diffeomorphism. Then f is ergodic.*

Proof By the Hopf Argument, it suffices to show that if ψ^s and ψ^u are L^2 functions with the following properties:

1. ψ^s is constant along leaves of $\mathcal{W}^s$,
2. ψ^u is constant along leaves of $\mathcal{W}^u$, and
3. $\psi^s = \psi^u$ a.e.,

then ψ^s (and so ψ^u as well) is constant a.e.

This is proved using the absolute continuity of $\mathcal{W}^u$ and $\mathcal{W}^s$. Since M is connected, one may argue this locally. Let G be the full measure set of $p \in M$ such that $\psi^s = \psi^u$. Absolute continuity of $\mathcal{W}^s$ (more precisely, consequence (AC1) of absolute continuity described above) implies that for almost every $p \in M$, G has full measure in $\mathcal{W}^s(p)$. Pick such a p . Then for almost every $q \in \mathcal{W}^s(p)$, $\psi^s(q) = \psi^u(p)$; defining G' to be the union over all $q \in \mathcal{W}^s(p) \cap G$ of $\mathcal{W}^u(q)$, one obtains that ψ^s is constant on $G \cap G'$. But now, since $\mathcal{W}^s(p) \cap G$ has full measure in $\mathcal{W}^s(p)$, the absolute continuity of $\mathcal{W}^u$ (consequence (AC2) above) implies that G' has full measure in a neighborhood of p . Hence ψ^s is a.e. constant in a neighborhood of p , completing the proof. $\square$

SRB Measures

In the absence of a smooth invariant measure, it is still possible for a map to have an invariant measure that behaves naturally with respect to volume. In computer simulations one observes such measures when one picks a point x at random and plots many iterates of x ; in many systems, the resulting picture is surprisingly insensitive to the initial choice of x . What appears to be happening in these systems is that the trajectory of almost every x in an open set U is converging to the support of a singular invariant probability measure μ . Furthermore, for any open set V , the proportion of forward iterates of x spent in V appears to converge to $\mu(V)$ as the number of iterates tends to 8.

In the 1960s and 70s, Sinai, Ruelle and Bowen rigorously established the existence of these

physically observable measures for hyperbolic attractors (Sinai 1972; Ruelle 1976; Bowen 1975). Such measures are now known as Sinai-Ruelle-Bowen (SRB) measures, and have been shown to exist for non-hyperbolic maps with some hyperbolic features. This subsection describes the construction of SRB measures for hyperbolic attractors.

An f -invariant probability measure μ is called an *SRB (or physical) measure* if there exists an open set $U \subset M$ containing the support of μ such that, for every continuous function $\phi : M \rightarrow \mathbb{R}$ and λ -a.e. $x \in U$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \phi(f^i(x)) = \int_M \phi d\mu.$$

The maximal open set U with this property is called the *basin* of f . To exclude the possibility that the SRB measure is supported on a periodic sink, one often adds the condition that at least one of the Lyapunov exponents of f with respect to μ is positive. Other definitions of SRB measure have been proposed (see (Young 2002)). Note that every ergodic absolutely continuous invariant measure with positive density in an open set is an SRB measure. Note also that an SRB measure for f is not in general an SRB measure for f^{-1} , unless f preserves an ergodic absolutely continuous invariant measure.

Every transitive Anosov diffeomorphism carries a unique SRB measure. To prove this, one defines a sequence of probability measures v_n on M as follows. Fix a point $p \in M$, and define v_0 to be the normalized restriction of Riemannian volume to a ball B^u in $\mathcal{W}^u(p)$. Set $v_n = \frac{1}{n} \sum_{i=1}^n f^i_* v_0$. Distortion estimates show that the density of v_n on its support inside $\mathcal{W}^u$ is bounded, above and below, independently of n . Passing to a subsequential weak* limit, one obtains a probability measure ν on M with bounded densities on $\mathcal{W}^u$ -leaves.

To show that ν is an SRB measure, choose a point $q \in M$ in the support of ν . Since ν has positive density on unstable manifolds, almost every point in a neighborhood of q in $\mathcal{W}^u(q)$ is a regular point for f (that is, a point where the

forward Birkhoff averages of every continuous function exist). A variation on the Hopf Argument, using the absolute continuity of $\mathcal{W}^s$, shows that ν is an ergodic SRB measure.

A similar argument shows that every transitive hyperbolic attractor admits an ergodic SRB measure. In fact this SRB measure has much stronger mixing properties, namely, it is Bernoulli. To prove this, one first constructs a Markov partition (Bowen 1970) conjugating the action of f to a Bernoulli shift. This map sends the SRB measure to a Gibbs state for a mixing Markov shift (see ► [Pressure and Equilibrium States in Ergodic Theory](#)). A result that subsumes all of the results in this section is:

Theorem 3 (Sinai, Ruelle, Bowen) *Let $\Lambda \subset M$ be a transitive hyperbolic attractor for a C^2 map $f: M \rightarrow M$. Then f has an ergodic SRB measure μ supported on Λ . Moreover: the disintegration of μ along unstable manifolds of Λ is equivalent to the induced Riemannian volume, the Lyapunov exponents of μ are all positive, and μ is Bernoulli.*

Beyond Uniform Hyperbolicity

The methods developed in the smooth ergodic theory of hyperbolic maps have been extended beyond the hyperbolic context. Two natural generalizations of hyperbolicity are:

- partial hyperbolicity, which requires uniform expansion of E^u and uniform contraction of E^s , but allows central directions at each point, in which the expansion and contraction is dominated by the behavior in the hyperbolic directions; and,
- nonuniform hyperbolicity, which requires hyperbolicity along almost every orbit, but allows the expansion of E^u and the contraction of E^s to weaken near the exceptional set where there is no hyperbolicity.

This section discusses both generalizations.

Partial Hyperbolicity

Brin and Pesin (1974) and independently Pugh and Shub (1972) first examined the ergodic properties of partially hyperbolic systems soon after the work of Anosov and Sinai on hyperbolic systems. One says that a diffeomorphism $f: M \rightarrow M$ of a compact manifold M is *partially hyperbolic* if there is a nontrivial, continuous splitting of the tangent bundle, $TM = E^s \oplus E^c \oplus E^u$, invariant under Df , such that E^s is uniformly contracted, E^u is uniformly expanded, and E^c is dominated, meaning that for some $n \geq 1$ and for all $x \in M$:

$$\|D_x f^n|_{E^s}\| < m(D_x f^n|_{E^c}) \leq \|D_x f^n|_{E^c}\| < m(D_x f^n|_{E^u}).$$

Partial hyperbolicity is a C^1 -open condition: any diffeomorphism sufficiently C^1 -close to a partially hyperbolic diffeomorphism is itself partially hyperbolic. For an extensive discussion of examples of partially hyperbolic dynamical systems, see the survey article (Burns et al. 2001) and the book (Pesin 2004). Among these examples are: the time-1 map of an Anosov flow, the frame flow for a compact manifold of negative sectional curvature, and many affine transformations of compact homogeneous spaces. All of these examples preserve the volume induced by a Riemannian metric on M .

As in the Anosov case, the stable and unstable bundles E^s and E^u of a partially hyperbolic diffeomorphism are tangent to foliations, again denoted by $\mathcal{W}^s$ and $\mathcal{W}^u$ respectively (Brin and Pesin 1974). Brin-Pesin and Pugh-Shub proved that these foliations are absolutely continuous.

A partially hyperbolic diffeomorphism $f: M \rightarrow M$ is *accessible* if any point in M can be reached from any other along an *su-path*, which is a concatenation of finitely many subpaths, each of which lies entirely in a single leaf of $\mathcal{W}^s$ or a single leaf of $\mathcal{W}^u$. Accessibility is a global, topological property of the foliations $\mathcal{W}^u$ and $\mathcal{W}^s$ that is the analogue of transversality of $\mathcal{W}^u$ and $\mathcal{W}^s$ for Anosov diffeomorphisms. In fact, the transversality of these foliations in the Anosov case immediately implies that every Anosov

diffeomorphism is accessible. Fundamental Principle # 4 suggests that accessibility might be related to ergodicity for conservative systems.

Conservative Partially Hyperbolic Diffeomorphisms

Motivated by a breakthrough result with Grayson (Grayson et al. 1994), Pugh and Shub conjectured that accessibility implies ergodicity, for a C^2 , partially hyperbolic conservative diffeomorphism (Pugh and Shub 1996). This conjecture has been proved under the hypothesis of center bunching (Burns and Wilkinson n.d.), which is a mild spectral condition on the restriction of Df to the center bundle E^c . Center bunching is satisfied by most examples of interest, including all partially hyperbolic diffeomorphisms with $\dim(E^c) = 1$. The proof in (Burns and Wilkinson n.d.) is a modification of the Hopf Argument using Lebesgue density points and a delicate analysis of the geometric and measure-theoretic properties of the stable and unstable foliations.

In the same article, Pugh and Shub also conjectured that accessibility is a widespread phenomenon, holding for an open and dense set (in the C^r topology) of partially hyperbolic diffeomorphisms. This conjecture has been proved completely for $r = 1$ (Dolgopyat and Wilkinson 2003), and for all r , with the additional assumption that the central bundle E^c is one dimensional (Rodríguez et al. 2008).

Together, these two conjectures imply the third, central conjecture: in (Pugh and Shub 1996):

Conjecture 1 (Pugh–Shub) *For any $r \geq 2$, the C^r , conservative partially hyperbolic diffeomorphisms contain a C^r open and dense set of ergodic diffeomorphisms.*

The validity of this conjecture in the absence of center bunching is currently an open question.

Dissipative Partially Hyperbolic Diffeomorphisms

There has been some progress in constructing SRB-type measures for dissipative partially

hyperbolic diffeomorphisms, but the theory is less developed than in the conservative case. Using the same construction as for Anosov diffeomorphisms, one can construct invariant probability measures that are smooth along the $\mathcal{W}^u$ foliation (Pesin and Sinai 1983). Such measures are referred to as u -Gibbs measures. Since the stable bundle E^s is not transverse to the unstable bundle E^u , the Anosov argument cannot be carried through to show that u -Gibbs measures are SRB measures.

Nonetheless, there are conditions that imply that a u -Gibbs measure is an SRB measure: for example, a u -Gibbs measure is SRB if it is the unique u -Gibbs measure (Dolgopyat 2004), if the bundle $E^s \oplus E^c$ is nonuniformly contracted (Bonatti and Viana 2000), or if the bundle $E^u \oplus E^c$ is nonuniformly expanded (Alves et al. 2000). The proofs of the latter two results use Pesin Theory, which is explained in the next subsection.

SRB measures have also been constructed in systems where E^c is nonuniformly hyperbolic (Burns et al. n.d.), and in (noninvertible) partially hyperbolic covering maps where E^c is 1-dimensional (Tsujii 2005). It is not known whether accessibility plays a role in the existence of SRB measures for dissipative, non-Anosov partially hyperbolic diffeomorphisms.

Nonuniform Hyperbolicity

The concept of Lyapunov exponents gives a natural way to extend the notion of hyperbolicity to systems that behave hyperbolically, but in a non-uniform manner. The fundamental principles described above, suitably modified, apply to these nonuniformly hyperbolic systems and allow for the development of a smooth ergodic theory for these systems. This program was initially proposed and carried out by Yakov Pesin in the 1970s (Pesin 1977) and has come to be known as *Pesin theory*.

Oseledec's Theorem implies that if a smooth map f satisfying the condition $m(D_x f) > 0$ preserves a probability measure ν , then for ν -a.e. $x \in M$ and every nonzero vector $v \in T_x M$, the limit

$$\lambda(x, v) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \log \|D_x f^i(v)\|$$

exists. The number $\lambda(x, v)$ is called the *Lyapunov exponent at x in the direction of v* . For each such x , there are finitely many possible values for the exponent $\lambda(x, v)$, and the function $x \mapsto \lambda(x, \cdot)$ is measurable. See the discussion of Oseledec's Theorem in ► [Ergodic Theorems](#).

Let f be a smooth map. An f -invariant probability measure μ is *hyperbolic* if the Lyapunov exponents of μ -a.e. point are all nonzero. Observe that any invariant measure of a hyperbolic map is a hyperbolic measure.

A conservative diffeomorphism $f: M \rightarrow M$ is *nonuniformly hyperbolic* if the invariant measure equivalent to volume is hyperbolic. The term “nonuniform” is a bit misleading, as uniformly hyperbolic conservative systems are also non-uniformly hyperbolic. Unlike uniform hyperbolicity, however, nonuniform hyperbolicity allows for the *possibility* of different strengths of hyperbolicity along different orbits.

Nonuniformly hyperbolic diffeomorphisms exist on all manifolds (Katok 1979; Dolgopyat and Pesin 2002), and there are C^1 -open sets of nonuniformly hyperbolic diffeomorphisms that are not Anosov diffeomorphisms (Shub and Wilkinson 2000). In general, it is a very difficult problem to establish whether a given map carries a hyperbolic measure that is nonsingular with respect to volume.

Hyperbolic Blocks

As mentioned above, the derivative of f along almost every orbit of a nonuniformly hyperbolic system looks like the derivative down the orbit of a uniformly hyperbolic system; the nonuniformity can be detected only by examining a positive measure set of orbits. Recall that Lusin's Theorem in measure theory states that every Borel measurable function is continuous on the complement of an arbitrarily small measure set. A sort of analogue of Lusin's theorem holds for nonuniformly hyperbolic maps: every C^2 , nonuniformly hyperbolic diffeomorphism is uniformly hyperbolic on a (noninvariant) compact set whose complement

has arbitrarily small measure. The precise formulation of this statement is omitted, but here are some of its salient features.

If μ is a hyperbolic measure for a C^2 diffeomorphism, then attached to μ -a.e. point $x \in M$ are transverse, smooth stable and unstable manifolds for f . The collection of all stable manifolds is called the *stable lamination* for f , and the collection of all unstable manifolds is called the *unstable lamination* for f . The stable lamination is invariant under f , meaning that f sends the stable manifold at x into the stable manifold for $f(x)$. The stable manifold through x is contracted uniformly by all positive iterates of f in a neighborhood of x . Analogous statements hold for the unstable manifold of x , with f replaced by f^{-1} .

The following quantities vary measurably in $x \in M$:

- the (inner) radii of the stable and unstable manifolds through x ,
- the angle between stable and unstable manifolds at x , and,
- the rates of contraction in these manifolds.

The stable and unstable laminations of a non-uniformly hyperbolic system are absolutely continuous. The precise definition of absolute continuity here is slightly different than in the uniformly and partially hyperbolic setting, but the consequences (AC1) and (AC2) of absolute continuity continue to hold.

Ergodic Properties of Nonuniformly Hyperbolic Diffeomorphisms

Since the stable and unstable laminations are absolutely continuous, the Hopf Argument can be applied in this setting to show:

Theorem 4 (Pesin) *Let f be C^2 , conservative and nonuniformly hyperbolic. Then there exists a (mod 0) partition $\mathcal{P}$ of M into countably many f -invariant sets of positive volume such that the restriction of f to each $P \in \mathcal{P}$ is ergodic.*

The proof of this theorem is also exposited in (Pugh and Shub 1989). The countable partition

can in examples be countably infinite; nonuniform hyperbolicity alone does not imply ergodicity.

The Dissipative Case

As mentioned above, establishing the existence of a nonsingular hyperbolic measure is a difficult problem in general. In systems with some global form of hyperbolicity, such as partial hyperbolicity, it is sometimes possible to “borrow” the expansion from the unstable direction and lend it to the central direction, via a small perturbation. Nonuniformly hyperbolic attractors have been constructed in this way (Viana 1997). This method is also behind the construction of a C^1 open set of nonuniformly hyperbolic diffeomorphisms in (Shub and Wilkinson 2000).

For a given system of interest, it is sometimes possible to prove that a given invariant measure is hyperbolic by establishing an *approximate* form of hyperbolicity. The idea, due to Wojtkowski and called the *cone method*, is to isolate a measurable bundle of cones in TM defined over the support of the measure, such that the cone at a point x is mapped by $D_x f$ into the cone at $f(x)$. Intersecting the images of these cones under all iterates of Df , one obtains an invariant subbundle of TM over the support of f that is nonuniformly expanded.

Lai-Sang Young has developed a very general method (Young 1998) for proving the existence of SRB measures with strong mixing properties in systems that display “some hyperbolicity”. The idea is to isolate a region X in the manifold where the first return map is hyperbolic and distortion estimates hold. If this can be done, then the map carries a mixing, hyperbolic SRB measure. The precise rate of mixing is determined by the properties of the return-time function to X ; the longer the return times, the slower the rate of mixing.

More results on the existence of hyperbolic measures are discussed in the next section.

An important subject in smooth ergodic theory is the relationship between entropy, Lyapunov exponents, and dimension of invariant measures of a smooth map. Significant results in this area include the Pesin entropy formula (Pesin 1976), the Ruelle entropy inequality (Ruelle 1978), the entropy-exponents-dimension formula

of Ledrappier and Young (1985a, 1985b), and the proof by Barreira-Pesin-Schmeling that hyperbolic measures have a well-defined dimension (Barreira et al. 1999). Hyperbolic Dynamical Systems contains a discussion of these results; see this entry there for further information.

The Presence of Critical Points and Other Singularities

Now for a discussion of the aforementioned technical difficulties that arise in the presence of singularities and critical points for the derivative.

Singularities, that is, points where Df (or even f) fails to be defined, arise naturally in the study of billiards and hard sphere gases. The first subsection discusses some progress made in smooth ergodic theory in the presence of singularities.

Critical points, that is, points where Df fails to be invertible, appear inescapably in the study of noninvertible maps. This type of complication already shows up for noninvertible maps in dimension 1, in the study of unimodal maps of the interval. The second subsection discusses the technique of parameter exclusion, developed by Jakobson, which allows for an ergodic analysis of a parametrized family of maps with criticalities.

The technical advances used to overcome these issues in the interval have turned out to have applications to dissipative, nonhyperbolic, diffeomorphisms in higher dimension, where the derivative is “nearly critical” in places. The last subsection describes extensions of the parameter exclusion technique to these near-critical maps.

Hyperbolic Billiards and Hard Sphere Gases

In the 1870s the physicist Ludwig Boltzmann hypothesized that in a mechanical system with many interacting particles, physical measurements (observables), averaged over time, will converge to their expected value as time approaches infinity. The underlying dynamical system in this statement is a Hamiltonian system with many degrees of freedom, and the “expected value” is with respect to Liouville measure. Loosely phrased in modern terms, Boltzmann’s hypothesis states that a generic Hamiltonian

system of this form will be ergodic on constant energy submanifolds. Reasoning that the time scales involved in measurement of an observable in such a system are much larger than the rate of evolution of the system, Boltzmann's hypothesis allowed him to assume that physical quantities associated to such a system behave like constants.

In 1963, Sinai revived and formalized this ergodic hypothesis, stating it in a concrete formulation known as the Boltzmann-Sinai Ergodic Hypothesis. In Sinai's formulation, the particles were replaced by N hard, elastic spheres, and to compactify the problem, he situated the spheres on a k -torus, $k = 2, 3$. The Boltzmann-Sinai Ergodic Hypothesis is the conjecture that the induced Hamiltonian system on the $2kN$ -dimensional configuration space is ergodic on constant energy manifolds, for any $N \geq 2$.

Sinai verified this conjecture for $N = 2$ by reducing the problem to a billiard map in the plane. As background for Sinai's result, a brief discussion of planar billiard maps follows.

Let $D \subset \mathbb{R}^s$ be a connected region whose boundary ∂D is a collection of closed, piecewise smooth simple curves the plane. The *billiard map* is a map defined (almost everywhere) on $\partial D \times [-\pi, \pi]$. To define this map, one identifies each point $(x, \theta) \in \partial D \times [-\pi, \pi]$ with an inward-pointing tangent vector at x in the plane, so that the normal vector to ∂D at x corresponds to the pair $(x, \pi/2)$. This can be done in a unique way on the smooth components of ∂D . Then $f(x, \theta)$ is obtained by following the ray originating at (x, θ) until it strikes the boundary ∂D for the first time at (x', θ') . Reflecting this vector about the normal at x' , define $f(x, \theta) = (x', \pi - \theta')$.

It is not hard to see that the billiard map is conservative. The billiard map is piecewise smooth, but not in general smooth: the degree of smoothness of f is one less than the degree of smoothness of ∂D . In addition to singularities arising from the corners of the table, there are singularities arising in the *second* derivative of f at the tangent vectors to the boundary.

In the billiards studied by Sinai, the boundary ∂D consists of a union of concave circular arcs and straight line segments. Similar billiards, but with convex circular arcs, were first studied by

Bunimovich (1974). Sinai and Bunimovich proved that these billiards are ergodic and non-uniformly hyperbolic. For the Boltzmann-Sinai problem with $N \geq 3$, the relevant associated dynamical system is a higher dimensional billiard table in Euclidean space, with circular arcs replaced by cylindrical boundary components.

In a planar billiard table with circular/flat boundary, the behavior of vectors encountering a flat segment of boundary is easily understood, as is the behavior of vectors meeting a circular segment in a neighborhood of the normal vector. If the billiard map is ergodic, however, every open set of vectors will meet the singularities in the table infinitely many times. To establish the non-uniform hyperbolicity of such billiard tables via cone-fieds, it is therefore necessary to understand precisely the fraction of time orbits spend near these singularities. Furthermore, to use the Hopf argument to establish ergodicity, one must avoid the singularities in the second derivative, where distortion estimates break down. The techniques for overcoming these obstacles involve imposing restrictions on the geometry of the table (even more so for higher dimensional tables), and are well beyond the scope of this paper.

The study of hyperbolic billiards and hard sphere gases has a long and involved history. See the articles (Szász 2000) and (Chernov and Markarian 2003) for a survey of some of the results and techniques in the area. A discussion of methods in singular smooth ergodic theory, with particular applications to the Lorentz attractor, can be found in (Araújo and Pacifico 2007). Another, more classical, reference is (Katok et al. 1986), which contains a formulation of properties on a critical set, due to Katok–Strelcyn, that are useful in establishing ergodicity of systems with singularities.

Interval Maps and Parameter Exclusion

The logistic family of maps $f_t: x \mapsto tx(1-x)$ defined on the interval $[0, 1]$ is very simple to define but exhibits an astonishing variety of dynamical features as the parameter t varies. For small positive values of t , almost every point in I is attracted under the map f_t to the sink at -1 . For values of $t > 4$, the map has a repelling hyperbolic

Cantor set. As the value of t increases between 0 and 4, the map f_t undergoes a cascade of period-doubling bifurcations, in which a periodic sink of period 2^n becomes repelling and a new sink of period 2^{n+1} is born. At the accumulation of period doubling at $t \approx 3.57$, a periodic point of period 3 appears, forcing the existence of periodic points of all periods. The dynamics of f_t for t close to 4 has been the subject of intense inquiry in the last 20 years.

The map f_t , for t close to 4, shares some of the features of the doubling map T_2 ; it is 2-to-1, except at the critical point $\frac{1}{2}$, and it is uniformly expanding in the complement of a neighborhood of this critical point. Because this neighborhood of the critical point is not invariant, however, the only invariant sets on which f_t is uniformly hyperbolic have measure zero. Furthermore, the second derivative of f_t vanishes at the critical point, making it impossible to control distortion for orbits that spend too much time near the critical point.

Despite these serious obstacles, Michael Jakobson (1981) found a method for constructing absolutely continuous invariant measures for maps in the logistic family. The method has come to be known as *parameter exclusion* and has seen application far beyond the logistic family. As with billiards, it is possible to formulate geometric conditions on the map f_t that control both expansion (hyperbolicity) and distortion on a positive measure set. As these conditions involve understanding infinitely many iterates of f_t , they are impossible to verify for a given parameter value t .

Using an inductive formulation of this condition, Jakobson showed that the set of parameters t near 4 that fail to satisfy the condition at iterate n have exponentially small measure (in n). He thereby showed that for a positive Lebesgue measure set of parameter values t , the map f_t has an absolutely continuous invariant measure (Jakobson 1981). This measure is ergodic (mixing) and has a positive Lyapunov exponent. The delicacy of Jakobson's approach is confirmed by the fact that for an open and dense set of parameter values, almost every orbit is attracted to a periodic sink, and so f_t has no absolutely

continuous invariant measure (Graczyk and Świątek 1997; Lyubich 1999). Jakobson's method applies not only to the logistic family but to a very general class of C^3 one-parameter families of maps on the interval.

Near-Critical Diffeomorphisms

Jakobson's method in one dimension proved to extend to certain highly dissipative diffeomorphisms. The seminal paper in this extension is due to Benedicks and Carleson; the method has since been extended in a series of papers (Mora and Viana 1993; Benedicks and Young 1993; Benedicks and Viana 2001) and has been formulated in an abstract setting (Wang and Young 2008).

This extension turns out to be highly non-trivial, but it is possible to describe informally the similarities between the logistic family and higher-dimensional "near critical" diffeomorphisms. The diffeomorphisms to which this method applies are crudely hyperbolic with a one dimensional unstable direction. Roughly this means that in some invariant region of the manifold, the image of a small ball under f will be stretched significantly in one direction and shrunk in all other directions. The directions of stretching and contraction are transverse in a large proportion of the invariant region, but there are isolated "near critical" subregions where expanding and contracting directions are nearly tangent.

The dynamics of such a diffeomorphism are very close to 1-dimensional if the contraction is strong enough, and the diffeomorphism resembles an interval map with isolated critical points, the critical points corresponding to the critical regions where stable and unstable directions are tangent.

An illustration of this type of dynamics is the Hénon family of maps $f_{a,b} : (x, y) \mapsto (1 - ax^2 + by, x)$, the original object of study in Benedicks-Carleson's work. When the parameter b is set to 0, the map $f_{a,b}$ is no longer a diffeomorphism, and indeed is precisely a projection composed with the logistic map. For small values of b and appropriate values of a , the Hénon map is strongly dissipative and displays the near critical behavior described in the previous paragraph. In analogy

to Jakobson's result, there is a positive measure set of parameters near $b = 0$ where $f_{a,b}$ has a mixing, hyperbolic SRB measure.

See (Viana and Luttsatto 2003) for a detailed exposition of the parameter exclusion method for Hénon-like maps.

Future Directions

In addition to the open problems discussed in the previous sections, there are several general questions and problems worth mentioning:

- What can be said about systems with everywhere vanishing Lyapunov exponents? Open sets of such systems exist in arbitrary dimension. Pesin theory carries into the non-uniformly hyperbolic setting the basic principles from uniformly hyperbolic theory (in particular, Fundamental Principles #3 and 4 above). To what extent do properties of isometric and unipotent systems (for example, Fundamental Principle #2) extend to conservative systems all of whose Lyapunov exponents vanish?
- Can one establish the existence of and analyze in general the conservative systems on surfaces that have two positive measure regimes: one where Lyapunov exponents vanish, and the other where they are nonzero? Such systems are conjectured exist in the presence of KAM phenomena surrounding elliptic periodic points.
- On a related note, how common are conservative systems whose Lyapunov exponents are nonvanishing on a positive measure set? See (Bochi and Viana 2004) for a discussion.
- Find a broad description of those dissipative systems that admit finitely (or countably) many physical measures. Are such systems dense among all dissipative systems, or possibly generic among a restricted class of systems? See (Shub and Wilkinson 2000; Palis 2005) for several questions and conjectures related to this problem.
- Extend the methods in the study of systems with singularities to other specific systems of

interest, including the infinite dimensional systems that arise in the study of partial differential equations.

- Carry the methods of smooth ergodic theory further into the study of smooth actions of discrete groups (other than the integers) on manifolds. When do such actions admit (possibly non-invariant) “physical” measures?

There are other interesting open areas of future inquiry, but this gives a good sample of the range of possibilities.

Bibliography

- Alves JF, Bonatti C, Viana M (2000) SRB measures for partially hyperbolic systems whose central direction is mostly expanding. *Invent Math* 140(2):351–398
- Anosov DV (1967) Geodesic flows on closed Riemann manifolds with negative curvature. *Proceedings of the Steklov Institute of Mathematics*, No. 90. Translated from the Russian by S. Feder. American Mathematical Society, Providence
- Anosov DV, Sinai JG (1967) Certain smooth ergodic systems. (Russian) *Uspehi Mat Nauk* 22 no. 5(137):107–172
- Araújo V, Pacifico MJ (2007) Three Dimensional Flows. *Publicações Matemáticas do IMPA*. [IMPA Mathematical Publications] 26° Col'quioio Brasileiro de Matemática. [26th Brazilian Mathematics Colloquium] Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro
- Arnol'd VI, Avez A (1986) *Ergodic problems of classical mechanics*. Translated from the French by Avez, A. Benjamin WA, Inc., New York/Amsterdam
- Baladi V (2000) Positive transfer operators and decay of correlations. *Advanced Series in Nonlinear Dynamics*, 16. World Scientific Publishing Co., Inc., River Edge
- Barreira L, Pesin Y, Schmeling J (1999) Dimension and product structure of hyperbolic measures. *Ann Math* 149(2):755–783
- Benedicks M, Carleson L (1991) The dynamics of the Hénon map. *Ann Math* 133(2):73–169
- Benedicks M, Viana M (2001) Solution of the basin problem for Hénon-like attractors. *Invent Math* 143(2):375–434
- Benedicks M, Young LS (1993) Sinai–Bowen–Ruelle measures for certain Hénon maps. *Invent Math* 112(3):541–576
- Bochi J, Viana M (2004) Lyapunov exponents: how frequently are dynamical systems hyperbolic? *Modern dynamical systems and applications*. Cambridge University Press, Cambridge, pp. 271–297
- Bonatti C, Viana M (2000) SRB measures for partially hyperbolic systems whose central direction is mostly contracting. *Israel J Math* 115:157–193

- Bonatti C, Diaz LJ, Viana M (2005) Dynamics beyond uniform hyperbolicity. A global geometric and probabilistic perspective. *Encyclopaedia of mathematical sciences*, 102. Mathematical Physics, III. Springer, Berlin
- Bowen R (1970) Markov partitions for Axiom A diffeomorphisms. *Am J Math* 92:725–747
- Bowen R (1975) Equilibrium states and the ergodic theory of Anosov diffeomorphisms. *Lecture Notes in Mathematics*, vol 470. Springer, Berlin/New York
- Brin MI, Pesin JB (1974) Partially hyperbolic dynamical systems. *Izv Akad Nauk SSSR Ser Mat* 38:170–212. (Russian)
- Brin M, Stuck G (2002) Introduction to dynamical systems. Cambridge University Press, Cambridge
- Bunimovic LA (1974) The ergodic properties of certain billiards. (Russian) *Funkcional Anal i Prilozhen* 8(3): 73–74
- Burns K, Wilkinson A, On the ergodicity of partially hyperbolic systems. *Ann of Math*. To appear
- Burns K, Pugh C, Shub M, Wilkinson A (2001) Recent results about stable ergodicity. Smooth ergodic theory and its applications. (Seattle, 1999), 327–366. *Proc Sympos Pure Math*, 69, Amer Math Soc, Providence
- Burns K, Dolgopyat D, Pesin Y, Pollicott, M. Stable ergodicity for partially hyperbolic attractors with negative central exponents. Preprint
- Chernov N, Markarian R (2003) Introduction to the ergodic theory of chaotic billiards. Second edition. *Publicações Matemáticas do IMPA*. IMPA Mathematical Publications 24° Colóquio Brasileiro de Matemática. 26th Brazilian Mathematics Colloquium Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro
- Cornfeld IP, Fomin SV, Sinai YG (1982) Ergodic theory. Translated from the Russian by A. B. Sosinskii. *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, vol 245. Springer, New York
- Dolgopyat D (2004) On differentiability of SRB states for partially hyperbolic systems. *Invent Math* 155(2): 389–449
- Dolgopyat D, Pesin Y (2002) Every compact manifold carries a completely hyperbolic diffeomorphism. *Ergod Theor Dynam Syst* 22(2):409–435
- Dolgopyat D, Wilkinson A (2003) Stable accessibility is C^1 dense. *Geometric methods in dynamics*. II. *Astérisque* No. 287
- Graczyk J, Swiatek G (1997) Generic hyperbolicity in the logistic family. *Ann Math* 146(2):1–52
- Grayson M, Pugh C, Shub M (1994) Stably ergodic diffeomorphisms. *Ann of Math* 140(2):295–329
- Hirsch MW (1979) Differential topology. *Graduate Texts in Mathematics*, No. 33. Springer, New York/Heidelberg
- Hirsch MW, Pugh CC, Shub M (1977) Invariant manifolds. *Lecture Notes in Mathematics*, vol 583. Springer, Berlin/New York
- Hopf E (1939) Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung. *Ber Verh Sachs Akad Wiss Leipzig* 91:261–304. (German)
- Jakobson MV (1981) Absolutely continuous invariant measures for one-parameter families of one-dimensional maps. *Commun Math Phys* 81(1):39–88
- Katok A (1979) Bernoulli diffeomorphisms on surfaces. *Ann Math* 110(2):529–547
- Katok A, Hasselblatt B (1995) Introduction to the modern theory of dynamical systems. With a supplementary chapter by Katok and Leonardo Mendoza. *Encyclopedia of Mathematics and its Applications*, vol 54. Cambridge University Press, Cambridge
- Katok A, Strelcyn JM, Ledrappier F, Przytycki F (1986) Invariant manifolds, entropy and billiards; smooth maps with singularities. *Lecture notes in mathematics*, 1222. Springer, Berlin
- Kifer Y (1986) Ergodic theory of random transformations. *Progress in probability and statistics*, vol 10. Birkhäuser, Boston
- Kifer Y (1988) Random perturbations of dynamical systems. *Progress in probability and statistics*, 16. Birkhäuser, Boston
- Ledrappier F, Young LS (1985a) The metric entropy of diffeomorphisms. I. Characterization of measures satisfying Pesin's entropy formula. *Ann Math* 122(2): 509–539
- Ledrappier F, Young LS (1985b) The metric entropy of diffeomorphisms. II. Relations between entropy, exponents and dimension. *Ann Math* 122(2):540–574
- Liu PD, Qian M (1995) Smooth ergodic theory of random dynamical systems, *Lecture notes in mathematics*, vol 1606. Springer, Berlin
- Lyubich M (1999) Feigenbaum–Coullet–Tresser universality and Milnor's hairiness conjecture. *Ann Math* 149(2):319–420
- Mañé R (1987) Ergodic theory and differentiable dynamics. Translated from the Portuguese by Silvio L. *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*, vol 8. Springer, Berlin
- Mora L, Viana M (1993) Abundance of strange attractors. *Acta Math* 171(1):1–71
- Palis J (2005) A global perspective for non-conservative dynamics. *Ann Inst H Poincaré Anal Non Linéaire* 22(4):485–507
- Pesin JB (1976) Characteristic Ljapunov exponents, and ergodic properties of smooth dynamical systems with invariant measure. *Dokl Akad Nauk SSSR* 226. (Russian)
- Pesin JB (1977) Characteristic Ljapunov exponents, and smooth ergodic theory. *Uspehi Mat Nauk* 32(196): 55–112. 287 (Russian)
- Pesin YB (2004) Lectures on partial hyperbolicity and stable ergodicity. Zurich lectures in advanced mathematics. European Mathematical Society (EMS), Zürich
- Pesin YB, Sinai YG (1983) Gibbs measures for partially hyperbolic attractors. *Ergod Theor Dynam Syst* 2(3–4):417–438

- Pugh C, Shub M (1972) Ergodicity of Anosov actions. *Invent Math* 15:1–23
- Pugh C, Shub M (1989) Ergodic attractors. *Trans Am Math Soc* 312(1):1–54
- Pugh C, Shub M (1996) Stable ergodicity and partial hyperbolicity. *International conference on dynamical systems (Montevideo, 1995)*, 182–187, Pitman Res Notes Math Ser, 362, Longman, Harlow
- Robinson C (1995) *Dynamical systems. Stability, symbolic dynamics, and chaos. Studies in Advanced Mathematics*. CRC Press, Boca Raton
- Rodríguez HA, Rodríguez HF, Ures R (2008) Partially hyperbolic systems with 1D-center bundle. *Invent Math* 172(2)
- Ruelle D (1976) A measure associated with axiom-A attractors. *Am J Math* 98(3):619–654
- Ruelle D (1978) An inequality for the entropy of differentiable maps. *Bol Soc Brasil Mat* 9(1):83–87
- Shub M, Wilkinson A (2000) Pathological foliations and removable zero exponents. *Invent Math* 139(3):495–508
- Sinai JG (1961) Geodesic flows on compact surfaces of negative curvature. *Dokl Akad Nauk SSSR* 136: 549–552 (Russian); translated as *Soviet Math Dokl* 2: 106–109
- Sinai JG (1972) Gibbs measures in ergodic theory. (Russian) *Uspehi Mat Nauk* 27 no. 4(166):21–64
- Szász D (2000) Boltzmann's ergodic hypothesis, a conjecture for centuries? Hard ball systems and the Lorentz gas, *Encycl Math Sci*, vol 101. Springer, Berlin, pp 421–448
- Tsujii M (2005) Physical measures for partially hyperbolic surface endomorphisms. *Acta Math* 194(1):37–132
- Viana M (1997) Multidimensional nonhyperbolic attractors. *Inst Hautes Études Sci Publ Math* 85: 63–96
- Viana M, Luttsatto S (2003) Exclusions of parameter values in Hénon-type systems. (Russian) *Uspekhi Mat Nauk* 58 (2003), no. 6(354), 3–44; translation in *Russian Math. Surveys* 58(6):1053–1092
- Wang QD, Young LS (2008) Toward a theory of rank one attractors. *Ann Math* 167(2)
- Young LS (1998) Statistical properties of dynamical systems with some hyperbolicity. *Ann Math* 147(2): 585–650
- Young LS (2002) What are SRB measures, and which dynamical systems have them? Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays. *J Stat Phys* 108(5–6):733–754

Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces

Krzysztof Frączek¹ and Corinna Ulcigrai²

¹Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland

²Institut für Mathematik, Universität Zürich, Zürich, Switzerland

Article Outline

Glossary

Definition of the Subject

Introduction

Examples

Locally Hamiltonian flows

Background and Tools

Invariant Measures and (Unique) Ergodicity

Behavior of Ergodic Averages

Mixing Properties

Spectral Properties

Disjointness Results

Open Directions

Bibliography

Glossary

Borel flow is a family $\varphi_{\mathbb{R}} = (\varphi_t)_{t \in \mathbb{R}}$ of Borel maps on a topological space X such that $(t, x) \mapsto \varphi_t x$ is also Borel, $\varphi_0 = Id_X$ and $\varphi_{t_1+t_2} = \varphi_{t_1} \circ \varphi_{t_2}$ for all $t_1, t_2 \in \mathbb{R}$. The flow preserves a Borel probability measure μ on X if $\mu(\varphi_t(A)) = \mu(A)$ for any Borel set A and any $t \in \mathbb{R}$. Then we say that $\varphi_{\mathbb{R}}$ is a **measure-preserving flow** of (X, μ) .

If S is compact smooth manifold and the map $(t, x) \mapsto \varphi_t x$ is smooth, then $\varphi_{\mathbb{R}}$ is called a **smooth flow**. If additionally S is a surface (dimension two manifold) and an invariant measure μ is smooth and positive (has a smooth positive density), then $\varphi_{\mathbb{R}}$ is called an **area-**

preserving smooth flow. Such flows are also called multi-valued or **locally Hamiltonian**.

We say that two flows $\varphi_{\mathbb{R}}$ on (X, μ) and $\phi_{\mathbb{R}}$ on (Y, ν) and are **isomorphic** as measure-preserving flows if there exists an isomorphism $\Phi : X \rightarrow Y$, i.e., a bimeasurable map which transports the measure μ to ν (i.e., $\mu(\Phi^{-1}(A)) = \nu(A)$ for any Borel set $A \subset Y$) and commutes with the dynamics, i.e., $\Phi(\varphi_t(x)) = \phi_t(\Phi(x))$ for μ -almost every (a.e.) $x \in X$. If additionally $\varphi_{\mathbb{R}}$, $\phi_{\mathbb{R}}$ and Φ are smooth, then the flows $\varphi_{\mathbb{R}}$ and $\phi_{\mathbb{R}}$ are **smoothly isomorphic**.

By a **joining** between two flows $\varphi_{\mathbb{R}}$ on (X, μ) and $\phi_{\mathbb{R}}$ on (Y, ν) we mean any probability $(\varphi_t \times \phi_t)_{t \in \mathbb{R}}$ -invariant measure on $X \times Y$ whose projections on X and Y are equal to μ and ν , respectively. Two flows $\varphi_{\mathbb{R}}$ and $\phi_{\mathbb{R}}$ are called **disjoint** (in the sense of Furstenberg) if their only common joining is the product joining $\mu \times \nu$.

A μ -preserving flow $\varphi_{\mathbb{R}}$ is **ergodic** if any invariant Borel set $A \subset X$ (i.e., $\mu(\varphi_t A \Delta A) = 0$ for all $t \in \mathbb{R}$) has measure zero or its complement $X \setminus A$ has measure zero.

A μ -preserving flow $\varphi_{\mathbb{R}}$ is **weakly mixing** if for any pair $f, g \in L_0^2(X, \mu)$ ($L_0^2(X, \mu)$ is the subspace of zero mean functions),

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \left| \int_X (f \circ \varphi_t) \cdot g \, d\mu \right| dt = 0.$$

A μ -preserving flow $\varphi_{\mathbb{R}}$ is **mixing** if for any pair $f, g \in L_0^2(X, \mu)$,

$$\lim_{t \rightarrow \infty} \int_X (f \circ \varphi_t) \cdot g \, d\mu = 0.$$

A μ -preserving flow $\varphi_{\mathbb{R}}$ is **rigid** if for any $f \in L^2(X, \mu)$,

$$\liminf_{t \rightarrow +\infty} \|f - f \circ \varphi_t\|_{L^2(X, \mu)} = 0.$$

A μ -preserving flow $\varphi_{\mathbb{R}}$ is **mildly mixing** if for any non-zero $f \in L^2(X, \mu)$,

$$\liminf_{t \rightarrow +\infty} \|f - f \circ \varphi_t\|_{L^2(X, \mu)} > 0.$$

A μ -preserving flow $\varphi_{\mathbb{R}}$ is **mixing of all orders** if for any $n \geq 2$ and any n -tuple $A_0, \dots, A_{n-1}$ of Borel sets,

$$\mu(A_0 \cap \varphi_{t_1}(A_1) \cap \dots \cap \varphi_{t_1 + \dots + t_{n-1}}(A_{n-1})) \xrightarrow{t_1, t_2, \dots, t_{n-1} \rightarrow \infty} \mu(A_0) \cdots \mu(A_{n-1}).$$

A Borel measure σ_f on $\mathbb{R}$ is the **spectral measure** of $f \in L^2(X, \mu)$ if

$$\int_{\mathbb{R}} e^{its} d\sigma_f(t) = \int_X (f \circ \varphi_s) \cdot \bar{f} d\mu \text{ for all } s \in \mathbb{R}.$$

The **spectral type** of a μ -preserving flow $\varphi_{\mathbb{R}}$ is an equivalence class of a Borel measure σ on $\mathbb{R}$ such that for every $f \in L^2_0(X, \mu)$, σ_f is absolutely continuous with respect to σ and there exists $f_0 \in L^2_0(X, \mu)$ such that $\sigma_{f_0} = \sigma$.

Definition of the Subject

We consider in this entry flows on closed, orientable smooth surfaces of genus $g \geq 1$ which preserve an invariant measure of full support and survey what is known of their ergodic, mixing, spectral, and joining properties. We focus in particular on compact surfaces and on two classes of area-preserving flows, namely translation flows on translation surfaces (and their Poincaré maps, interval exchange transformations (IETs)) and smooth area-preserving flows known as locally Hamiltonian or multi-valued Hamiltonian flows.

Introduction

Flows on surfaces are one of the most basic and fundamental examples of dynamical systems, studied since the birth of dynamical systems as a research field with the work of Poincaré (1987) at the end of the nineteenth century. Many models of systems of physical origin are described by flows on surfaces, starting from celestial mechanics, polygonal billiard dynamics, and up to solid-

state physics, or statistical mechanics models. Despite being low-dimensional systems of zero topological entropy, they present a rich display of fine chaotic properties.

We define below two classes of area-preserving flows, linear flows on translation surfaces and locally Hamiltonian flows, whose ergodic and spectral properties have been the object of intense research activity in the last decades. Area-preserving flows on surfaces also provide one of the fundamental classes of *parabolic*, or *slowly chaotic*, dynamical systems (see also the survey Ulcigrai (2021)). Contrary to hyperbolic systems, which display *sensitive dependence* on initial conditions (i.e., divergence in time of nearby initial conditions, the so-called *butterfly effect*) which happens (infinitesimally) at *exponential speed*, parabolic systems display a slow form of sensitive divergence, with speed of divergence which is sub-exponential (and actually *polynomial* or sub-polynomial in all known examples).

Brief historical remarks Early results on flows on the ergodic theory of flows on surfaces date back to the 1970s (two articles Aranson and Grines (1973) and Katok (1973) can be seen as an impetus for development). Interval exchange transformations, discrete maps which naturally appear as Poincaré sections of flows on surfaces, were introduced and studied earlier (see, e.g., the work by Oseledec (1966)) independently, as an interesting example to study spectral aspects in ergodic theory. The study of the ergodic theory of linear flows on translation surfaces has taken off in the 1980s in connection with the study of interval exchange transformations and billiards in rational polygons, with seminal works by Masur and Veech laying the grounds for the connection with Teichmüller dynamics, a research area which has since then benefited from the contribution of several Fields medalists (including Avila, Kontsevich, McMullen, Mirzakhani, and Yoccoz).

The study of locally Hamiltonian flows has seen a revival of interest in the 1990s, with the seminal work of Zorich (1984) in connection with the Novikov problem. Novikov indeed underlined the interest in the study of multi-valued Hamiltonians and the associated flows (Novikov 1982) in

connection with problems arising in solid-state physics as well as in pseudo-periodic topology (see, e.g., the survey by Zorich (1999)). Indeed, Novikov (1982) and his school in the 1990s advocated the study of locally Hamiltonian flows as model to describe the motion of an electron in a metal under a magnetic field in the semi-classical approximation (the surface appears here as Fermi energy level surface). Novikov made some conjectures (known as *Novikov problem*) on the asymptotic behavior of trajectories of electrons. At the same time, Arnold (1991) made a conjecture on mixing for the flows we call today *Arnold flows*. This conjecture has been the motivation for a lot of the work on the mixing properties of locally Hamiltonian flows.

The current century has seen a lot of advances in our understanding of the chaotic properties of smooth area-preserving flows (a class which includes locally Hamiltonian flows), in particular exploiting Teichmüller dynamics tools, but also under the influence of the work of Marina Ratner in homogeneous dynamics. The study of linear flows on infinite translation surfaces has only begun in the last decade, in particular motivated by the study of the periodic Ehrenfest model and it is still a widely open research direction.

Examples

Linear flows on the torus The basic example of an area-preserving flow is the linear flow on the torus $\mathbb{T}^2 := \mathbb{R}^2 / \mathbb{Z}^2$ (see Fig. 1) given by solutions $(x(t), y(t)) \in \mathbb{T}^2$ to

$$(x'(t), y'(t)) = (\cos \theta, \sin \theta) \quad \theta \in S^1, \quad (1)$$

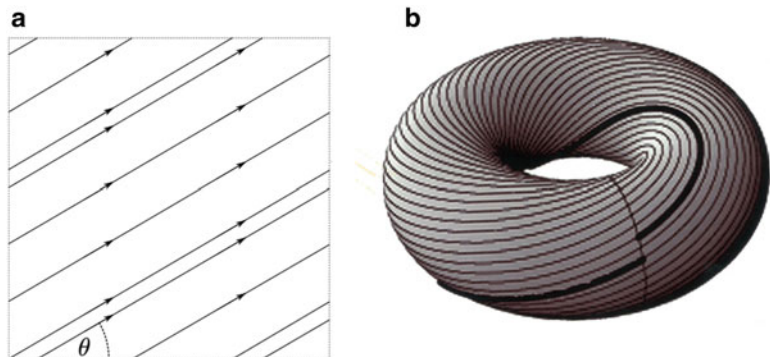
which moves points at unit speed along (the image by the projection $p : \mathbb{R}^2 \rightarrow \mathbb{R}^2 / \mathbb{Z}^2$ of) Euclidean lines in direction θ (i.e., lines making an angle θ with the horizontal axes); see Fig. 1a.

Linear flows on translation surface Linear flows (also called *translation flows*) can be defined more in general on *translation surfaces*, namely surfaces which are locally Euclidean outside a finite number of conical singularities. A translation surface can be defined as a quotient

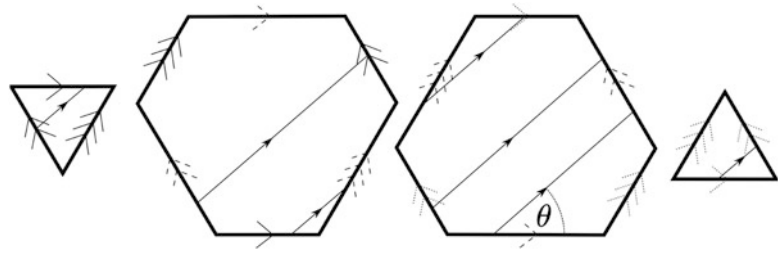
$$S := P_1 \cup P_2 \cup \dots \cup P_n / \sim$$

where $P_i \subset \mathbb{R}^2$ for $1 \leq i \leq n$ are disjoint polygons in $\mathbb{R}^2$, with clockwise oriented boundary, with the property that their edges can be paired into couples (e, e') where e and e' are parallel and isometric and have opposite orientation (see an example in Fig. 2) and the equivalence relation $\sim$ identifies the edges of the pair (e, e') by the (unique) translation of $\mathbb{R}^2$ which maps e to e' . The request that e and e' have opposite orientations is needed to guarantee that, after gluing, one obtains a translation surface, in which all sides are identified by translations, and not a *half-translation* surface, where identifications by a central symmetry composed with a translation are allowed. Translation surfaces can be equivalently defined as the datum of an Abelian differential on a Riemann surface, while half-translation surfaces correspond to quadratic differentials.

Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 1 A linear flow on the torus $\mathbb{T}^2 := \mathbb{R}^2 / \mathbb{Z}^2$.



**Ergodic and Spectral
Theory of
Area-Preserving Flows
on Surfaces,**
Fig. 2 Translation surface



The resulting space S is a compact, oriented topological surface which is endowed with a flat, Euclidean metric outside a finite set $\Sigma = \Sigma(S)$ the (image after gluings) of the set of vertices of the polygons. Points in Σ are known as *singularities*; each of these points has a neighborhood where the metric has a conical singularity with a total cone angle $2\pi k$, $k \in \mathbb{N}$ (i.e., it is locally of the form $ds^2 = dr^2 + (k r d\theta)^2$). Translation surfaces can be equivalently defined as a pair (X, ω) where X is a compact Riemann surface and ω is an Abelian differential; we refer to the surveys (Masur and Tabachnikov 2002; Yoccoz 2010; Forni and Matheus 2014).

On a translation surface S , the notion of a direction $\theta \in S^1$ (in particular the notion of horizontal and vertical direction) is well defined not only locally, in each polygon, but globally (since the identifications are performed by translations). Thus, for each $\theta \in S^1$, we can globally define a flow $\varphi_{\mathbb{R}}^{\theta} := (\varphi_t^{\theta})_{t \in \mathbb{R}}$ given locally by solutions to (1). As in the case of the torus, $\varphi_{\mathbb{R}}^{\theta}$ moves points along the (quotient to S) of lines in direction θ with unit speed; see Fig. 2. Notice that at a singularity $v \in \Sigma$ with cone angle $2\pi k$, there is not a unique but k lines in direction θ and v is a saddle point of the linear flow $\varphi_{\mathbb{R}}^{\theta}$ with $2k$ -prongs, of the type shown in Fig. 4c for $k = 3$. Notice furthermore that these flows preserve a Euclidean area, but are discontinuous flows, since singularities are reached in finite time.

Locally Hamiltonian flows

Smooth (in particular continuous) flows on S preserving a smooth measure can be defined as

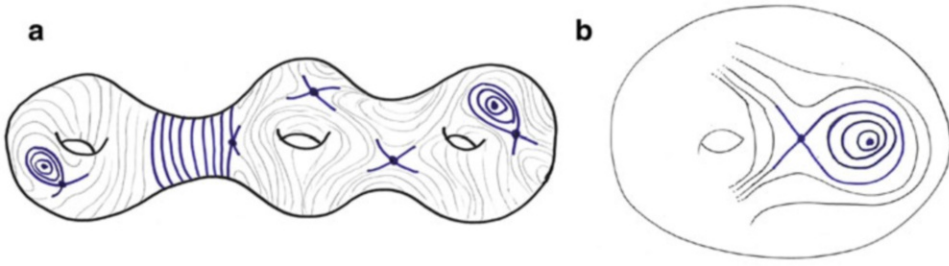
follows. Let ω be a fixed smooth area form (locally given in coordinates (x, y) by $f(x, y)dx \wedge dy$ where f is a smooth positive function). Thus, equivalently, the pair (S, ω) is a two-dimensional symplectic manifold. We consider smooth flows $\varphi_{\mathbb{R}}$ on S which preserve a measure μ given by integrating a smooth density with respect to ω . We assume that the area is normalized so that $\mu(S) = 1$. It turns out that such smooth area-preserving flows on S are in one-to-one correspondence with smooth closed real-valued differential 1-forms as follows. Given a smooth, closed, real-valued differential 1-form η , let X be the vector field determined by $\eta = i_X \omega$ where i_X denotes the contraction operator, i.e., $i_X \omega = \omega(X, \cdot)$ and consider the flow $\varphi_{\mathbb{R}}$ on S given by X . Since η is closed, the transformations φ_t , $t \in \mathbb{R}$, are area-preserving. Conversely, every smooth area-preserving flow can be obtained in this way.

The flow $\varphi_{\mathbb{R}}$ is known as the *multi-valued Hamiltonian* flow associated to η . Indeed, the flow $\varphi_{\mathbb{R}}$ is *locally Hamiltonian*, i.e., *locally* one can find coordinates (x, y) on S in which $\varphi_{\mathbb{R}}$ is given by the solution to the equations

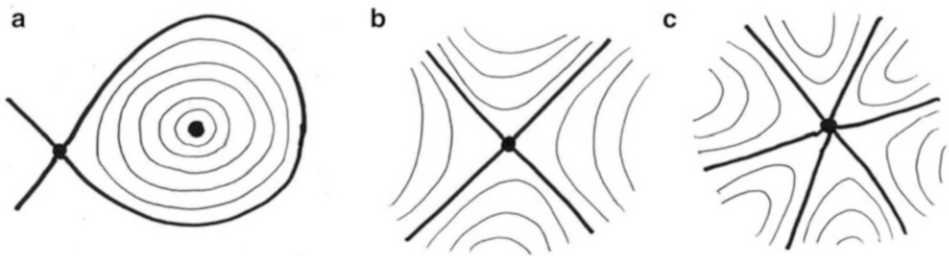
$$\begin{cases} \dot{x} = \partial H / \partial y, \\ \dot{y} = -\partial H / \partial x \end{cases}$$

for some smooth real-valued Hamiltonian function H . A *global* Hamiltonian H cannot be in general defined (see Nikolaev and Zhuzhoma (1999), Section 1.3.4), but one can think of $\varphi_{\mathbb{R}}$ as globally given by a *multi-valued* Hamiltonian function.

Locally Hamiltonian flows necessarily have fixed points or *singularities*. Singularities, as shown in Fig. 4, can be either centers (Fig. 4a),



Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 3 Examples of flows

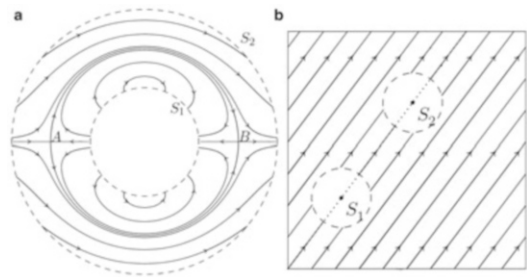


Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 4 Type of singularities

simple saddles (Fig. 4b) or multi-saddles (i.e., saddles with $2k$ pronges, $k \geq 2$; see Fig. 4c for $k = 3$). Examples of flow trajectories are shown in Fig. 3. For $g = 1$, i.e., on a torus, if there is a singularity, then there has to be another one.

Arnold flows The simplest examples of locally Hamiltonian flows with singularities on a torus, i.e., flows with one center and one simple saddle (see Fig. 3b), were studied by Arnold (1991) and are nowadays often called *Arnold flows*. Any *Arnold flow* is the restriction of such locally Hamiltonian flow to a minimal component obtained by removing the center and the disk filled by periodic orbits around it (called *island*), which, as Arnold shows in (Arnold 1991), is always bounded by a saddle loop.

Blokhin examples A class of special examples in higher genus which has been studied much earlier than the general case are Blokhin flows; see Blohin (1972). These are flows on surfaces of higher genus $g \geq 2$ obtained essentially by gluing together via a suitable surgery flows on genus one components. More precisely, Blokhin's



Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 5 Blokhin flows

construction is based on a linear ergodic flow on the torus (Fig. 5b) and a flow on the annulus, which has two simple saddles connected by separatrices (Fig. 5a). Two identical circles are cut in the torus so that their centers lie in the same orbit of the line flow. The annulus is glued into the cut-out circles so that the separatrices connect to the orbit passing through the center of the removed circles. The resulting flow on the genus two surface has a special representation over an irrational rotation. We can perform the same pasting procedure several times increasing the genus of the surface.

Time-changes New flows can be obtained by perturbation starting from the examples in the previous section. The simplest perturbation is a *time-change*, also known as a *time-reparametrization*, which produces flows that move points along the *same orbits*, but with different *speed*. The flow $\tilde{\phi}_{\mathbb{R}}$ is a *time-change* (or *reparametrization*) of a flow $\phi_{\mathbb{R}}$ on S if there exists a measurable function $\tau : S \times \mathbb{R} \rightarrow \mathbb{R}$ such that for every $x \in S$ the map $t \mapsto \tau(x, t)$ is continuous and strictly increasing, and for all $x \in S$ and $t \in \mathbb{R}$ we have $\tilde{\phi}_{\tau(x,t)}(x) = \phi_t(x)$. Since $\tilde{\phi}_{\mathbb{R}}$ is assumed to be a flow, the function τ is an *additive cocycle* over the flow $\phi_{\mathbb{R}}$, that is, it satisfies the cocycle identity:

$$\tau(x, s + t) = \tau(\phi_s(x), t) + \tau(x, s)$$

for all $x \in S$ and all $s, t \in \mathbb{R}$. If $\phi_{\mathbb{R}}$ is a smooth flow, we will say that $\tilde{\phi}_{\mathbb{R}}$ is a *smooth* reparametrization if the cocycle τ is a smooth function such that the maps $\tau(x, \cdot)$ are homeomorphisms. Even these simple perturbations can produce a genuinely *new* flow, i.e., a flow which is *not* measurably conjugated to the unperturbed flow (see below). More drastic (non-smooth) time-changes, when τ is allowed to have *singularities* where it blows up, can introduce singularities in the time-changed flow. The simplest type of singularity is a *stopping point* (also known as *fake saddle point*), i.e., a fixed point which has a neighborhood foliated by flow trajectories.

There is an obvious way to produce measurably (or even smoothly) conjugated flows using reparametrizations, associated to solutions to the so-called *cohomological equation* as follows. Two additive cocycles $\tau_1(x, t)$, $\tau_2(x, t)$ are said to be measurably (respectively smoothly) *cohomologous* if their difference $\tau_1 - \tau_2$ is a measurable (respectively smooth) coboundary, i.e., if there exists a measurable (respectively smooth) function $u : M \rightarrow \mathbb{R}$, called the *transfer function*, such that

$$\tau_1(x, t) - \tau_2(x, t) = u(x) - u \circ \phi_t(x) \quad (2)$$

for all $(x, t) \in S \times \mathbb{R}$. An elementary, but fundamental, result establishes that time-changes given

by cohomologous cocycles are isomorphic (see, e.g., Avila et al. (2021a), Lemma 2.1, or Katok (2001), §9, for related results). The time-change $\tilde{\phi}_{\mathbb{R}}$ given by the cocycle τ is *trivial* if (2) with $\tau_1 = \tau$ and $\tau_2 = id$ admits a solution. In this case, $\tilde{\phi}_{\mathbb{R}}$ and $\phi_{\mathbb{R}}$ are conjugated by the conjugacy $x \mapsto \tilde{\phi}_{u(x)}x$ and hence the time-change is isomorphic to the original flow. Equation (2) is an example of a *cohomological equation*. Thus, trivial time-changes are described by solutions to the so-called *cohomological eq.* A key feature of area-preserving flows is the existence of *obstructions* (invariant distributions) to solve cohomological equations; see Forni (1997). As a consequence, among smooth time-changes, *smoothly* trivial time-changes are *rare* (i.e., form a finite or countable codimension subspace) and time-changes can display essentially different spectral and ergodic properties compared to the original flow. Nevertheless, certain ergodic properties, like *ergodicity* and *cohomological properties*, which only depend on the orbit structure and hence are independent of the time-change, persist in a time-change.

Background and Tools

Minimality and minimal components From the topological dynamics point of view, one is interested in the qualitative behavior of every trajectory. An example where all orbits can be understood are linear flows on tori (e.g., given by (1)), which satisfy a well-known dichotomy: either all orbits are *periodic*, namely there exists a $t_0 > 0$ such that $\varphi_{t_0+t}(x) = \varphi_t(x)$ for every $x \in S$ and $t \in \mathbb{R}$ (this happens exactly when the slope θ in (1) is rational), or $\varphi_{\mathbb{R}}$ is *minimal*, i.e., every (forward) trajectory $\{\varphi_t(x), t \geq 0\}$, for any $x \in S$, is dense.

In presence of fixed points (see Fig. 4), which are unavoidable in higher genus, the definition of minimality should be adjusted as follows. We say that the trajectory of $x \in S$ is *regular* if the whole orbit $\{\varphi_t(x), t \in \mathbb{R}\}$ is well defined and the limits $\lim_{t \rightarrow \pm \infty} \varphi_t(x)$ do not exist. A flow $\varphi_{\mathbb{R}} : S \rightarrow S$ is called *quasi-minimal* (or simply *minimal*, by

abusing the terminology) if every *regular* trajectory is dense. If $\varphi_{\mathbb{R}}$ has a fixed point that is a center (Fig. 4a), then since a neighborhood of the center is foliated by periodic orbits, $\varphi_{\mathbb{R}}$ cannot be (quasi-)minimal.

Let us say that a segment of a trajectory, of the form $\{\varphi_t(x), a < t < b\}$ (where possibly $a = -\infty$ or $b = +\infty$) is a *saddle connection* if it starts and ends in a saddle point (when a or b is ∞ , by this we mean that $\lim_{t \rightarrow \pm\infty} \varphi_t(x)$ are saddle points). A saddle connection is called a *saddle loops* if the initial and final saddles are the same (see Fig. 6a).

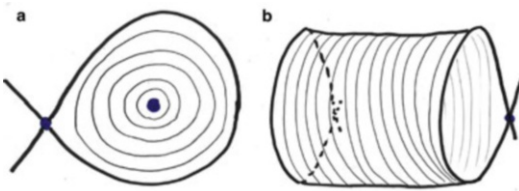
A classical result on flows on surfaces (see, e.g., Yoccoz (2010)) is that if $\varphi_{\mathbb{R}}$ has no saddle connections, then it is (quasi-)minimal. The corresponding result at the level of IETs, proved by Keane (1975), is that an IET T with an irreducible combinatorial datum π such that the orbits discontinuities of T are all infinite and distinct (a condition called I.D.O.C. by Keane (1975) and nowadays known as *Keane's condition* in the literature; see, e.g., Yoccoz (2010)) is minimal.

Notice that if $\varphi_{\mathbb{R}}$ has a fixed point which is a center, one can show that the center is contained in a disk filled with closed (i.e., periodic) trajectories and bounded by a saddle loop or a union of saddle

connections, called an *island* of periodic orbits; see Fig. 6a. Hence, in presence of centers, the flow $\varphi_{\mathbb{R}}$ is never minimal (since orbits in the complement of the island avoid the island and hence cannot be dense).

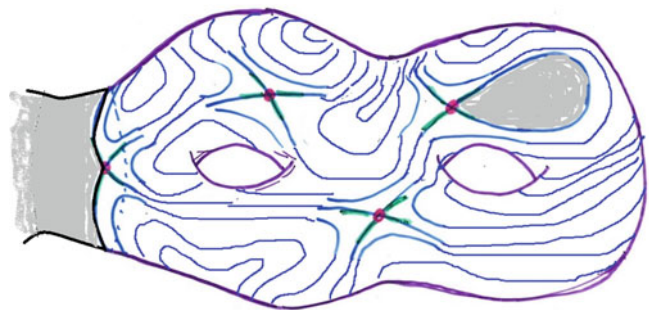
Mayer (1943), Levitt (1983), and Zorich (1999) proved independently that each smooth area-preserving flow can be decomposed into a finite number of subsurfaces with boundary S_i , $i = 1, \dots, N$ such that for each i the restriction of $\varphi_{\mathbb{R}}$ to S_i is a *periodic component*, i.e., the interior of S_i is foliated into closed orbits of $\varphi_{\mathbb{R}}$, or S_i is such that the restriction of $\varphi_{\mathbb{R}}$ to S_i is (quasi-)minimal; the latter are called *minimal components*. Periodic components are either islands (as in Fig. 6a) or cylinders filled by periodic orbits and bounded by saddle loops, as in Fig. 6b. Minimal components (see an example in Fig. 7a), by topological reasons, cannot be more than g (the genus of S). The flows in Fig. 3, for example, can be decomposed, in the case of 3a, into three periodic components, two islands and one cylinder filled by closed orbits, and two minimal components (one of genus one and one of genus two), while, in the case of the flow on the torus in Fig. 3b, there is one island and one minimal component (the so-called Arnold flow).

One can show (see below) that minimal components of a locally Hamiltonian flow (and in particular minimal such flows, for which S is in itself a minimal component) are *time-reparametrization* (see the previous section) of linear flows on translation surfaces (although they time-change in this case is singular), so in particular, they have the same orbits and the same topological behavior as linear flows (see, e.g.,



Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 6 Periodic components

Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 7 A minimal component



Zorich (1999). This was in part one of the original motivations (in addition to the unfolding of rational billiards in the West) that sparked the interest of mathematicians such as Zorich in the ergodic theory of *linear flows*.

Interval exchange maps and Poincaré sections A central idea introduced by Poincaré was that the study of a surface flow can be often *reduced* to the study of a one-dimensional discrete dynamical system, by taking what we nowadays call a *Poincaré section* and considering the *Poincaré first return map* of the flow to the section (when and where it is defined).

In genus one, if we consider the linear flow (1) on $\mathbb{T}^2$ and take the horizontal side $[0, 1] \times \{0\} \subset \mathbb{T}^2$, the Poincaré map is the (rigid) rotation $R_\alpha : [0, 1] \rightarrow [0, 1]$ given by $x \mapsto R_\alpha(x) = x + \alpha \bmod 1$ (where $\alpha = \cot \theta$, see (1)). More generally, if we start from a flow $\varphi_{\mathbb{R}} := (\varphi_t)_{t \in \mathbb{R}}$ on a torus, i.e., on a compact, orientable surface S of genus one, and assume that it does not have fixed points or closed orbits (or more generally Reeb components; see Nikolaev and Zhuzhoma (1999)), there is a (global) section given by a closed transverse curve and the Poincaré first return map to it is a diffeomorphism $f : S^1 \rightarrow S^1$ of the circle $S^1 \cong \mathbb{R}/\mathbb{Z}$.

Interval exchange transformations (IETs) As in the case of genus one, an essential tool to study a higher genus flow is to consider a (local) *transversal* $I \subset S$ to the flow and the *Poincaré first return map* T of the flow on I (when it is defined, for example, almost everywhere when the flow preserves a finite measure with full support; see more generally (Nikolaev and Zhuzhoma 1999)).

Consider first a linear flow on a translation surface. In this case, Poincaré maps are piecewise isometries of an interval, known as *interval exchange transformations* (or for short, IETs): a one-to-one map $T : I \rightarrow I$ of $I = [0, 1)$ is a (standard) interval exchange transformations of $d \geq 2$ intervals, or *d-IET* for short, if one can partition I into intervals $I_1, \dots, I_d$ so that the restriction T_i of T to I_i , for each $1 \leq i \leq d$, is a translation. Thus, for any I_i there exists

$\delta_i \in [-1, 1]$ such that $T(x) = x + \delta_i$ for any $x \in I_i$. Notice that a rotation R_α can be seen as a 2-IET, which exchanges the intervals $[0, 1-\alpha)$ and $[1-\alpha, 1)$. Thus IETs provide a natural generalization of circle rotations and play for linear flows in higher genus an analogous role to rotations for linear flows on tori. In general, the images $T(I_1), \dots, T(I_d)$ are *exchanged intervals* which form a new partition of $[0, 1)$. A *d-IET* is completely determined by two data, namely a *combinatorial datum* π which determines the *order* of the exchanged intervals in $[0, 1)$ (this can be a permutation of $\{1, \dots, d\}$ or a *pair* of permutations of an alphabet of cardinality d – see Yoccoz (2010) – a convention which is often useful to study fine chaotic properties of IETs) and a length vector $\lambda = (\lambda_1, \dots, \lambda_d)$ which belongs to the simplex

$$\Delta_d := \left\{ \lambda \in \mathbb{R}_+^d, \lambda_i \geq 0 \text{ for all } 1 \leq i \leq d, \quad (3) \right. \\ \left. \sum_{i=1}^d \lambda_i = 1 \right\}.$$

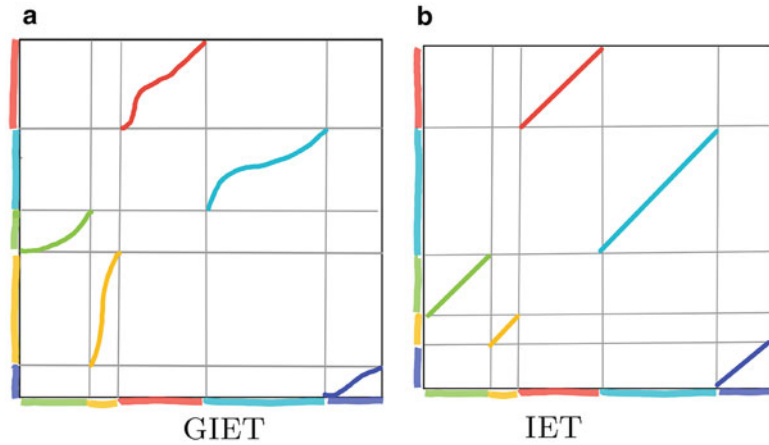
Then the corresponding IET $T = T_{(\pi, \lambda)}$ exchanges the intervals

$$I_i = [l_i, r_i) = \left[\sum_{0 \leq j < i} \lambda_j, \sum_{0 \leq j \leq i} \lambda_j \right) \text{ for } i = 1, \dots, d$$

according to the permutation π (Fig. 8b). We say that the combinatorial datum is *irreducible* if the indexes $\{1, \dots, d\}$ of subintervals cannot be decomposed into two subsets $\{1, \dots, k\}$ and $\{k+1, d\}$ with $1 \leq k < d$ which are exchanged by themselves.

More generally, given any (smooth) flow $\varphi_{\mathbb{R}}$ on S , not necessarily preserving a smooth invariant measure, first return maps $T : I \rightarrow I$ are one-to-one *piecewise diffeomorphisms* known as *generalized interval exchange transformations*: a map $T : I \rightarrow I$ is a generalized interval exchange transformations or, for short, a GIET (Fig. 8a), if one can partition I into intervals $I_1, \dots, I_d$ (finitely many since we are assuming that $\varphi_{\mathbb{R}}$ has finitely many fixed points) so that the restriction T_i of T to

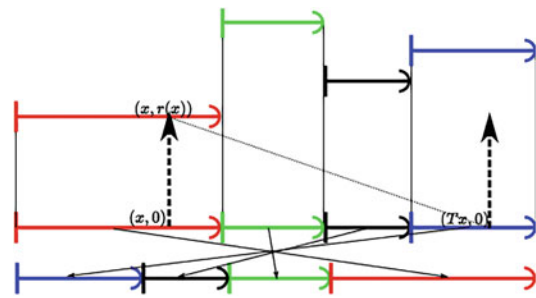
Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 8 A generalized IET (GIET) and a (standard) IET with $d = 4$



I_i , for each $1 \leq i \leq d$, is a (smooth) diffeomorphism onto its image which extends to a diffeo of the closure $\bar{I}_i$ (see, e.g., Marmi et al. (2012)), as shown in Fig. 8. Thus, IETs are a special case of GIETs (from which the use of the adjective *generalized*) and provide natural *linear model* of a GIET. Under the assumption that $\varphi_{\mathbb{R}}$ preserves a smooth (non-atomic) finite invariant measure, one can always choose suitable coordinates in which the Poincaré section is a standard IET, i.e., there is a (smooth) reparametrization of the section such that the GIET Poincaré section becomes a standard IET. In general, the *linearization problem* of characterizing the GIETs which are (smoothly) conjugated to standard IET is an open research area; see Ulcigrai (2022) for a survey.

Special flows representations Linear flows and minimal components of area-preserving flows can be conveniently described using *special flows*. These representations provide a concrete tool to work with them and study their fine ergodic and spectral properties. Given a Poincaré map of a flow to a section and the additional information of the return time to the section, special flows allow reconstructing (a measurable conjugated version of) the flow. We recall the construction for our special setting.

Let $T : I \rightarrow I$ be an (ergodic) IET and let $r : I \rightarrow \mathbb{R}_{>0} \cup \{+\infty\}$ be an integrable function such that $\underline{r} = \inf_{x \in I} r(x) > 0$. The *special flow*



Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 9 Special representation of a linear flow over IET with $d = 4$

over T under the *roof function* r is the flow $\phi_{\mathbb{R}}^{T,r} := (T_t^r)_{t \in \mathbb{R}}$ acting on

$$I' := \{(x, s) \in I \times \mathbb{R} : 0 \leq s < r(x)\},$$

so that $\phi_t^{T,r}(x, s) = (x, s + t - r^{(n)}(x))$, where $r^{(n)}(x)$ denotes the Birkhoff sums cocycle, i.e., the additive cocycle defined by

$$r^{(n)}(x) := \sum_{0 \leq k < n} r(T^k x) \quad \text{if } n \geq 0,$$

$$r^{(n)}(x) := - \sum_{n \leq k < 0} r(T^k(x)) \quad \text{if } n < 0,$$

associated to r and n is the unique integer number with $r^{(n)}(x) \leq s + t < r^{(n+1)}(x)$. It describes the motion of a point in $(x, s) \in I' \subset I \times \mathbb{R}$ along vertical trajectories, modulo the identification of each point $(x, r(x))$, $x \in I$, with the point $(Tx, 0)$; see Fig. 9.

Special representations of linear flows Linear flows on translation surfaces can be represented as special flows over IETs, under roof functions which are *piecewise constant*: more precisely, the return time $r : I \rightarrow \mathbb{R}_{>0}$ of a linear flow on a translation surface S to a transverse section $I \subset S$ is constant on each continuity interval I_i of the IET $T = T_{(\pi, \lambda)} : I \rightarrow I$ induced as Poincaré map; see Fig. 9. Then the roof function can be identified with a vector $r \in \mathbb{R}_{>0}^d$. As noted by Veech (1982b), the vector r is in a 2 g -dimensional subspace $H(\pi) \subset \mathbb{R}^d$, where g is the genus of the surface S .

Smooth time-changes of the linear flow produce flows which can still be represented as special flows over T , under a roof r which is piecewise smooth on each I_i and extends smoothly to its closure; see Fig. 10a.

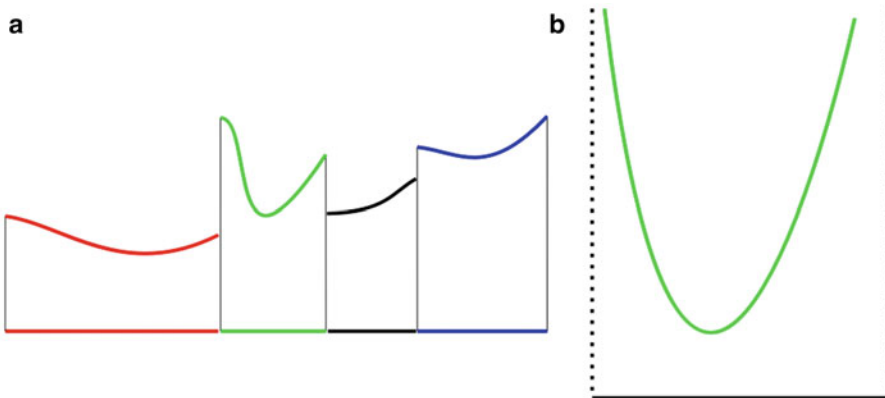
Special representation of Arnold flows It is well known that (minimal components of) locally Hamiltonian flows can be represented as special flows over IETs, but in this case the function r is *singular*, i.e., blows up at (some) endpoints of continuity intervals (see, e.g., Ulcigrai (2011), Ravotti (2017), Conze and Frączek (2011), and Frączek and Ulcigrai (2012)). The nature of the singularities of r turns out to depend crucially on the nature and type of fixed points of the surface flow. The first remark on the importance of such representation and the nature of the singularities for the study of chaotic properties (in particular

mixing) was Arnold. In Arnold (1991) he states that what we nowadays call *Arnold flow*, i.e., a minimal component of a locally Hamiltonian flow on a torus with a center and simple saddle, can be represented as a special flow over a rotation $R_\alpha : [0, 1] \rightarrow [0, 1]$ under a roof function of the form

$$r(x) = C |\log(x)| + 2C |\log(1-x)| \quad (4)$$

for some $C > 0$. This function is said to have *logarithmic singularities* at the endpoints of $[0, 1]$; see Fig. 10b. The singularities are furthermore called *asymmetric*, since the constants $2C \neq C$ are different. The singularities at $x = 0$ and $x = 1$ correspond to the trajectory which is a separatrix, i.e., ends in the saddle. Because of the nature of the Hamiltonian local parametrization of a saddle, it takes infinite time to reach the saddle and the motion along trajectories which are close to the separatrix is slowed down logarithmically; see, e.g., (Frączek and Ulcigrai 2012), [Appendix A] for a calculation. The fact that the constant $2C$ is exactly the *double* of the other constant C can be explained because trajectories on one side of the saddle loop pass *twice* near the saddle, while on the other side only once.

Special representation of locally Hamiltonian flows with simple saddles Singularities which are *simple saddles* (standard saddles with



Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 10 Special representation of flows

4-prongs, as in Fig. 4b) produce singularities of the roof function as several singularities which, like in the case of Arnold flows, have logarithmic nature in the following sense. We say that a function $r : I \rightarrow \mathbb{R}$ for an IET $T_{(\pi, \lambda)}$ has *logarithmic singularities* if there exist constants $C_i^+, C_i^- \in \mathbb{R}, i = 1, \dots, d$, and a function g_r absolutely continuous on the interior of each interval $I_i, i = 1, \dots, d$ (i.e., with the notation that we will introduce later, a function $g_r \in \text{AC}(\sqcup_{i=1}^d I_i)$) such that

$$\begin{aligned} r(x) = & - \sum_{i=1}^d C_i^+ \log(|I| \{(x - l_i)/|I|\}) \\ & - \sum_{i=1}^d C_i^- \log(|I| \{(r_i - x)/|I|\}) + g_r(x) \end{aligned} \quad (5)$$

See Fig. 11 for an example. Consider now either a minimal locally Hamiltonian flow $\varphi_{\mathbb{R}}$ on S or the restriction of a locally Hamiltonian flow on S to a minimal component $S' \subset S$. Let η be the associated closed 1-form and assume that η is Morse (the corresponding local Hamiltonian is a Morse function), so all saddles of the flow $\varphi_{\mathbb{R}}$ are simple. Then $\varphi_{\mathbb{R}}$ can be shown to be (measure theoretically) *isomorphic* to a *special flow* $T' : I' \rightarrow I'$ over an interval exchange transformation $T : I \rightarrow I$ of $d \geq 1$ intervals and under a roof r with logarithmic

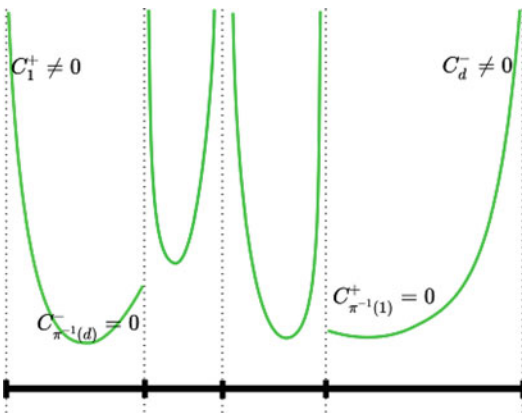
singularities. The number of exchanged intervals is $d = 2g + s - 1$ in the case when $\varphi_{\mathbb{R}}$ is minimal and s is the number of simple saddles, or, for a minimal component $S', d = 2g' + s' - 1$, where g' is the genus of S' and s' is the number of (simple) saddles in the closure of S' .

We say that the logarithmic singularities are of *geometric type* if at least one among C_d^- and $C_{\pi^{-1}(d)}^-$ is zero and at least one among C_1^+ or $C_{\pi^{-1}(1)}^+$ is zero (as shown in the examples in Fig. 11). We denote by $\text{LG}(\sqcup_{i=1}^d I_i)$ the space of functions with logarithmic singularities of geometric type. One can furthermore show that the roof functions arising from suitably chosen special representations (namely when the section is *standard*, i.e., both endpoints belong to saddle separatrices) are of geometric type. This notion plays a crucial role in some results on locally Hamiltonian flows; see, e.g., (Frączek and Ulcigrai 2012, 2021).

Symmetric and asymmetric logarithmic singularities When all the saddles are simple, and hence the roof has logarithmic singularities, it is crucial, for understanding finer chaotic features, to distinguish between *symmetric* or *asymmetric* singularities, in the following sense. Let $\text{LSG}(\sqcup_{i=1}^d I_i)$ be the subspace of functions with geometric type logarithmic singularities $\text{LG}(\sqcup_{i=1}^d I_i)$ which in addition satisfy the symmetry condition

$$\sum_{i=1}^d C_i^- - \sum_{i=1}^d C_i^+ = 0. \quad (6)$$

When the locally Hamiltonian flow is minimal on S and has *only* simple saddles, the representation leads to a *symmetric* roof (see, e.g., Kočergin (1976) and Ulcigrai (2011)). On the other hand, when the flow $\varphi_{\mathbb{R}}$ has centers or more in general saddle loops homologous to zero (i.e., which disconnect the surface, as the saddle loop surrounding a periodic island in Fig. 4a), minimal components lead to representations with *asymmetric* roofs. The prime example of an asymmetric special flow is the roof (4) of the special



Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 11 Example of roof with geometric logarithmic singularities

representation of an Arnold flow, while minimal flows with only nondegenerate saddles give roofs with symmetric logarithmic singularities; cf. also Blokhin flows.

Degenerate saddles and power-singularities -

Consider now *degenerate* saddles (i.e., saddles with $2k$ pronges, $k > 2$; see Fig. 4c for $k = 3$). These saddles produce stronger, *power-type* singularities of the roof function; see Kočergin (1975). We say that l_i or r_i is a power-like singularity of power-type $0 < \gamma < 1$ if there exists a constant $C_i^\pm$ such that

$$\begin{aligned} \lim_{x \rightarrow l_i^+} (x - l_i)^\gamma r(x - l_i) &= C_i^+; \\ \lim_{x \rightarrow r_i^-} (r_i - x)^\gamma r(r_i - x) &= C_i^-. \end{aligned} \quad (7)$$

Thus r behaves like the functions $C_i^\pm/x^\gamma$ in a neighborhood of l_i or r_i . As a degenerate special case, a *stopping point* (fake saddle) introduces a single symmetric power-singularity of the roof function.

Time-changes through special representations

One can show that two special flows over the same transformation T under two different roof functions are one a time-change of the other. Thus, comparing the special representations results recalled in this section for linear flows and minimal (components of) locally Hamiltonian flows, one can show that the latter are time-changes of translation flows via a *singular* reparametrization (which amount to changing an r which is piecewise constant to one which has singularities).

Von Neumann flows Another class of special flows over IETs which has been studied in the literature are so-called *von Neumann flows*. The name is used to denote special flows over rotations or IETs under a function r which is *piecewise absolutely continuous* (and continuous on each continuity interval of the IET) with non-zero sum of jumps, i.e., $\int r'(x)dx \neq 0$; see Frączek and Lemańczyk (2009a). These were first studied by von Neumann (1932) in the case of rotations in the

base, hence the name. Von Neumann flows also admit an interpretation as special representations of area-preserving surface flows. Consider a locally Hamiltonian flow which is non-minimal: assume that there are saddle connections which separate the surface into subsurfaces with boundary, each of which is a periodic or a minimal component. Note that due to the presence of fixed points, the transition times for the flow along the saddle connections are infinite. Using a reparametrization of the flow that results in significant acceleration around fixed points, described in Frączek and Lemańczyk (2009b), we obtain a flow for which all transition times along saddle connections are finite. Then, each minimal component of such reparametrization admits a representation as a von Neumann flow. Furthermore, the sum of jumps of the roof is related to the sum of transition times (summed according to orientation) along the saddle connections forming the boundary of the minimal component (cf. Conze and Frączek (2011)).

Open sets and typical behavior In order to describe the dynamical behavior of a *generic* or *typical* (area-preserving) surface flow, one can introduce the following topologies and measures or measure classes (see below) on the space of IETs, linear flows, and locally Hamiltonian flows and distinguish between open sets with different typical dynamical properties.

Almost every IET and almost every linear flow

Translation surfaces and their linear flows, as well as their Poincaré sections, IETs, are described by finite-dimensional spaces. We say that a result holds for *almost every* d -IET on a unit interval if it holds for all irreducible data π on d -symbols and almost every length vector λ in the simplex Δ_d ; see (3). Since linear flows can be represented as special flows over IETs under piecewise constant roof functions, it is sufficient to add $2g$ data (related to the values of the return time on each continuity interval). A result holds for *almost every* linear flow if it holds for a.e. IET in the base and a.e. choice of $r \in \mathbb{R}_{>0}^d \cap H(\pi)$. Translation surfaces (up to cut and paste

equivalence) can be locally parametrized by *period coordinates*, namely by the (holonomy) vectors associated to pairs of identified sides (see also, e.g., Zorich (2006), Yoccoz (2010), and Viana (n.d.)) for a more intrinsic definition using relative homology).

These finite-dimensional coordinates allow defining the so-called Masur-Veech measure. Perhaps the simplest way to define the Masur-Veech measure is to consider a presentation of a translation surface S as a polygon with $2N$ sides with pairs of parallel, congruent sides $(v_1, v'_1), \dots, (v_N, v'_N)$ identified by gluings. Then the vectors $(v_1, \dots, v_N) \in \mathbb{R}^{2N}$ give local coordinates for an open set U of translation surfaces around S and the Masur-Veech measure is just the Lebesgue measure on $U \subset \mathbb{R}^{2N}$. More formally, the Masur-Veech measure is the pull-back of the Lebesgue on $\mathbb{R}^{2n}$ by the period map which maps S to $\mathbb{R}^{2n}$ where n is the dimension of the relative homology group $H_1(S, \Sigma, \mathbb{R})$. Since the vertical flow on a translation surface can be represented as a special flow over an IET under a piecewise constant roof, a result which holds for the vertical flow on a.e. translation surface with respect to this measure holds for a.e. linear flow in the sense above. Notice though that to reconstruct the geometry of the translation surface, in addition to the IET and the roof function, one needs also additional data (see, e.g., the zippered rectangle construction introduced by Veech (1982b); see also (Yoccoz 2010)).

Genericity notions for locally Hamiltonian flows Let us denote by $\mathcal{F}$ the set of smooth closed 1-forms on S (i.e., locally Hamiltonian flows) with isolated zeros. One can define a *topology* on $\mathcal{F}$ by considering perturbations of closed smooth 1-forms by (small) closed smooth 1-forms: we say that a smooth, closed 1-form η' is an ϵ -perturbation of the smooth closed 1-forms η if for any $x \in S$ there exist coordinates on a simply connected neighborhood U of x , such that $\eta|_U = dH$ and $(\eta' - \eta)|_U = dh$ where $\|h\|_{C^\infty} < \epsilon \|H\|_{C^\infty}$. We then say that a condition is *generic* (in the sense of Baire) if it holds for flows described by an open and dense set of forms with respect to this topology. The subset

$\mathcal{M} \subset \mathcal{F}$ of *Morse closed 1-forms*, i.e., forms which are locally the differential of a *Morse function*, is *open and dense* in $\mathcal{F}$ with respect to this topology (see, e.g., Ravotti (2017)). Thus, a *generic* locally Hamiltonian flow (in the sense of Baire) has only *nondegenerate* fixed points, i.e., centers and simple saddles (see Fig. 4a, b), as opposed to degenerate *multi-saddles* (as in Fig. 4c).

Katok fundamental class and measure class Let us fix an open set $\mathcal{A}_{s,c}$ of closed 1-forms with c centers and s (simple) saddles. A *measure-theoretical notion of typical* on $\mathcal{A}_{s,c}$ can be defined on each $\mathcal{A}_{s,c}$ by using the *Katok fundamental class*, introduced by Katok (who was motivated, as he explains in Katok (1973), by previous work of Aranson and Grines, who had previously introduced in Aranson and Grines (1973) a homotopic topological invariant for flows on oriented surfaces) in Katok (1973). Let Σ be the set of zeros (fixed points) of η . Let $\gamma_1, \dots, \gamma_n$ be a base of the relative homology $H_1(S, \Sigma, \mathbb{R})$, where $n = 2g + s + c - 1$. The image of η by the period map Per is $\text{Per}(\eta) = \left(\int_{\gamma_1} \eta, \dots, \int_{\gamma_n} \eta \right) \in \mathbb{R}^n$. The map Per is well defined in a neighborhood of η in $\mathcal{A}_{s,c}$ and one can show (using Moser's so-called *homotopy trick*; see Katok (1973) and also Prop. 2.7 in Ravotti (2017)) that it is a complete isotopy invariant. The pull-back $\text{Per}_* \text{Leb}$ of the Lebesgue measure class (i.e., of the equivalence class of measures which have the same sets of measure zero as the Lebesgue measure) by the period map gives the desired measure class on closed 1-forms in $\mathcal{A}_{s,c}$. When we use the expression *typical* below (or *typical* in $\mathcal{U}_{\min}$ or $\mathcal{U}_{-\min}$) we mean full measure in each $\mathcal{A}_{s,c}$ with respect to this measure class on each $\mathcal{A}_{s,c}$ (or on each open subset of $\mathcal{A}_{s,c}$ contained in the union $\mathcal{U}_{\min}$ or $\mathcal{U}_{-\min}$). If one can show that a result holds for almost linear flow (or for the vertical flow on almost every translation surface with respect to the Masur-Veech measure), then it follows that it holds for a full measure set of locally Hamiltonian flows with respect to the Katok fundamental class (see, e.g., Ravotti (2017)).

The open sets $\mathcal{U}_{\min}$ and $\mathcal{U}_{\neg\min}$ To classify chaotic behavior in locally Hamiltonian flows it is crucial to distinguish between two (complementary, up to measure zero) open sets. Remark that if the flow $\varphi_{\mathbb{R}}$ given by a closed 1-form η has a *saddle loop homologous to zero* (i.e., the saddle loop is a *separating curve* on the surface), then the saddle loop is persistent under small perturbations (see Sect. 2.1 in Zorich (1999) or Lemma 2.4 in Ravotti (2017)). In particular, the set of locally Hamiltonian flows which have at least one saddle loop is open and gives the set denoted $\mathcal{U}_{\neg\min}$ above. Flows in the open set $\mathcal{U}_{\neg\min}$ admit a non-trivial decomposition into periodic components and minimal components.

The second open set, which we call $\mathcal{U}_{\min}$, is given by the interior (which one can show to be non-empty) of the complement of $\mathcal{U}_{\neg\min}$, i.e., the set of locally Hamiltonian flows without saddle loops homologous to zero. One can show that saddle loops non-homologous to zero (and saddle connections) vanish after arbitrarily small perturbations and neither the set of 1-forms with saddle loops non-homologous to zero (or saddle connections) nor its complement is open (see Ravotti (2017) for details).

Full measure of minimality In view of Keane's criterium for minimality (Keane 1975), one can show that the set of non-minimal IETs has measure zero and, furthermore, Hausdorff codimension 1. In view of the definition of measure class on locally Hamiltonian flows, we therefore deduce that:

Theorem 1 *In $\mathcal{U}_{\min}$, the typical flow is minimal (in particular there are no centers and there is a unique minimal component) and the typical flow on $\mathcal{U}_{\neg\min}$ is minimal when restricted to each component which is not a periodic component (bounded by saddle loops homologous to zero).*

Special flows representations in $\mathcal{U}_{\min}$ and $\mathcal{U}_{\neg\min}$

The typical locally Hamiltonian flow in $\mathcal{U}_{\min}$ admits a representation as a special flow over a (minimal) IET under a roof with logarithmic singularities (see (5)) and the roof satisfies the

symmetry condition (6). For a typical flow in $\mathcal{U}_{\neg\min}$, the restriction of the flow on each minimal component admits a representation as a special flow over a minimal IET under a roof with *asymmetric* logarithmic singularities (like in the case of the Arnold flow; see (4)). See, for example, (Ravotti 2017).

Invariant Measures and (Unique) Ergodicity

Let us discuss the existence and finiteness of ergodic measures, (unique) ergodicity, and non-uniquely ergodic examples.

Invariant measures: existence and finiteness Given a (topological) flow $\varphi_{\mathbb{R}}$ on S , since we are assuming that S is compact, the existence of a finite (probability) invariant measure is guaranteed by the Krylov-Bogolybov theorem. Katok shows in (Katok 1973) that any topologically transitive surface flow with non-degenerate (Morse) saddles has a probability invariant measures with non-trivial support and no atoms, positive on open sets.

When S is endowed with a reference smooth area form ω and $\varphi_{\mathbb{R}}$ is assumed to be smooth (at least outside the singularity set), it is natural to ask about the existence of a (finite) invariant measure which is absolutely continuous with respect to ω and in particular of measures which have a smooth (or at least differentiable) density. If the flow can be *linearized*, i.e., conjugated to a linear flow, via a differentiable or smooth conjugacy, such measures can be obtained by pull-back of the Lebesgue measure via the conjugacy.

Linearization questions and results Questions about linearization and regularity of the conjugacy are hard and largely open. Marmi et al. showed in Marmi et al. (2012) that for almost every linear flow $\varphi_{\mathbb{R}}$ on S , for any $r \geq 2$, among C^{r+3} -perturbations supported outside the singularity set $\Sigma \subset S$, those which are *linearizable* with a C^r conjugacy form a submanifold of finite

codimension $(g - 1)(2r + 1) + s$, where g denotes the genus of the surface and s denotes the number of singularities. In Marmi et al. (2012), it is conjectured that for $r = 1$ those which are C^1 -linearizable form a submanifold of codimension $3g + s - 3$. This was proved by Ghazouani (2021) for a measure-zero special case (hyperbolic periodic-type IETs).

In very recent work, Ghazouani and Ulcigrai (2021) have proved a rigidity theorem for foliations associated to smooth flows with Morse saddles on surfaces (at the moment only of genus 2): for almost all (a.a.) smooth flows with respect to the Katok fundamental class, if the corresponding foliation is topologically conjugate to an orientable measured foliation of that fundamental class, then it is C^1 conjugate. This result confirms (in genus 2) another conjecture of Marmi et al. (2012).

Bounds on the number of ergodic invariant measures It is a general fact that finite (resp. probability) invariant measures form a cone (resp. a simplex) generated by ergodic measures. In our setting, the number of independent ergodic probability measures turns out to be *finite*: indeed, as observed by Oseledec (1966), the number of independent ergodic invariant measures for a d -interval exchange transformation T is bounded d (see, e.g., Yoccoz (2010) for a proof). In fact, Oseledec estimated the maximal spectral multiplicity of any d -IET from above by d . This can be seen by showing that a d -IET has rank d .

Let us assume that $\varphi_{\mathbb{R}}$ preserves a smooth invariant probability measure and is minimal (or restrict it to a minimal component $S' \subset S$). By considering the (minimal) interval exchange transformations which appear as Poincaré sections of $\varphi_{\mathbb{R}}$ to sections (after linearizing the GIET exploiting the invariant measure induced by the smooth invariant measure for $\varphi_{\mathbb{R}}$), one can deduce from the IETs upper bound that the number of independent ergodic invariant probability measures for $\varphi_{\mathbb{R}}$ is also at most d , where d is the number of exchanged intervals and hence can be chosen (by taking the section to be normal, i.e., with endpoints on separatrices) to be $d = 2g + s - 1$ where g is the genus of S and s is the cardinality of fixed points.

A sharper bound was proved by Katok (1973), who showed that the number of invariant measures is at most $g \leq [d/2]$, where g is the genus of S (or of the minimal component S'). The proof of the dimension bound uses the idea of Katok fundamental class and the symplectic structure on the absolute homology $H_1(S, \mathbb{R})$ given by the intersection form; see also Forni (2002) and McMullen (2020).

First non-uniquely ergodic examples Katok's upper bound is optimal, as was shown by Sataev (1975), who constructed examples of flows on S with a cone of invariant measure of arbitrary dimension less or equal to the genus g . An example of a minimal non-uniquely ergodic IET with $d = 4$ was discovered by Keane (1977) as a counterexample of an early form of the so-called *Keane conjecture* (see below) which stated that all minimal interval exchanges are (uniquely) ergodic. Earlier examples also appeared in the work of Katok and Stepin (1966) and an example with $d = 5$ was built by Keynes and Newton (1976) building on Veech (1969). A modern revisitation of Keane's counterexample (which was produced with an ad-hoc self-induction procedure) can be obtained using the tools given by Rauzy-Veech induction (which were developed only later). Exploiting this point of view, in the lecture notes (Yoccoz 2010), examples of minimal non-uniquely ergodic IETs are produced in any Rauzy class.

Another question which has been investigated recently for non-uniquely ergodic IETs is *genericity*, i.e., the existence of an (a fortiori non-uniquely ergodic) IET T and a point $x \in I$ whose orbit under T equidistributes for a measure that is not ergodic. A d -IET with these properties has been constructed by Chaika and Masur (2015).

Keane conjecture and typical unique ergodicity Keane conjectured in Keane (1977) that almost every IET is uniquely ergodic. *Keane conjecture* (as it became known) was proved in 1982 at the same time by Masur (1982) and Veech (1982b). Both proofs exploit renormalization and introduced seminal ideas and techniques and are considered early milestones of the successful

application of Teichmüller dynamics to the study of IETs and translation surfaces; see, e.g., the surveys Chaika and Weiss (2022) or Zorich (2006). While Veech works directly on IETs developing an induction algorithm known as *Rauzy-Veech induction*, Masur shows that almost every linear flow (and hence a.e. IET) is uniquely ergodic by using geometric arguments and the Teichmüller geodesic flow on the moduli space of translation surfaces. Later in 1992, Masur proves in Masur (1992) the celebrated *Masur's criterion* for unique ergodicity, which shows that unique ergodicity follows from the assumption that the Teichmüller geodesic starting from the given translation surface is not divergent in moduli space.

Ergodicity everywhere, in almost every direction An important strengthening of Masur and Veech results was later proved by Kerckhoff et al. (1986):

Theorem 2 *For every translation surface, the linear flow in almost every direction is uniquely ergodic.*

This result covers in particular billiards in rational polygons, which can be *unfolded* (according to the Zemljakov & Katok construction in Zemljakov and Katok (1975); see also Fox and Kershner (1936)) to linear flows on (a measure zero set of) translation surfaces.

Consequences for locally Hamiltonian flows The first examples of smooth flows on surfaces which are *ergodic* were the so-called *Blokhin examples* (see Blohin (1972)). There are measure zero examples since they are essentially glued out of genus one subsurfaces. Since ergodicity of a (minimal component of a) locally Hamiltonian flow is equivalent to ergodicity of (and hence any) interval exchange transformation which appears as the Poincaré map, result on ergodicity of typical locally Hamiltonian flow can be deduced from the proof of Keane's conjecture by Masur and Veech and the relation between the two notions of full measure. We obtain:

Theorem 3 *Almost every locally Hamiltonian flow in $\mathcal{U}_{\min}$ is not only minimal, but also ergodic. For almost every flow in $\mathcal{U}_{\neg\min}$, the restriction of the flow on each minimal component is ergodic.*

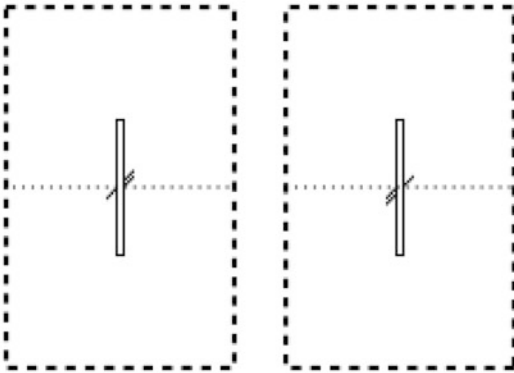
Due to the presence of saddle fixed points, locally Hamiltonian flows (their restrictions to minimal components) on higher genus surfaces are never uniquely ergodic.

Hausdorff dimension of non-uniquely ergodic directions Both Masur (1992) and Masur and Smillie (1991) study the exceptional set $\mathcal{N}\mathcal{U}(S)$ of directions on a given translation surface S which are minimal but fail to be uniquely ergodic. They show that not only this set has measure zero by the proof of Keane's conjecture, but it is also small in the Hausdorff dimension point of view:

Theorem 4 *For every translation surface S , the Hausdorff dimension of the set $\mathcal{N}\mathcal{U}(S)$ of non-uniquely ergodic directions does not exceed $1/2$. Furthermore, for any connected component C in the moduli space of translation surfaces with $g > 1$ there is a constant $c > 0$, called the Masur-Smillie constant of component C , such that for almost every translation surface $S \in C$ the Hausdorff dimension of $\mathcal{N}\mathcal{U}(S)$ is exactly c .*

Sporadic examples of non-uniquely ergodic interval were found, as already mentioned, by several authors before the proof of Keane's conjecture Keane (1977); Keynes and Newton (1976).

The slit-tori $g = 2$ example A much studied example of a translation surface where one can produce and study the exceptional set $\mathcal{N}\mathcal{U}$ of non-uniquely ergodic directions is the surface of genus two obtained by gluing two identical flat tori along a slit; see Fig. 12. This surface, introduced by Masur and Smillie, can be used to give a geometric presentation related to previous work of Veech (1969) on skew products over rotations. Cheung showed in Cheung (2003) that the upper bound of $1/2$ for the Hausdorff dimension of $\mathcal{N}\mathcal{U}$ is achieved in these examples. A full dichotomy was later proved by Cheung et al. (2011) that



Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 12 Example a slit torus

showed that the Hausdorff dimension $HD(\mathcal{N}\mathcal{U})$ is either 0 or $1/2$ (according to a Diophantine-like property on the length of the slit). Athreya and Chaika (2015) showed that in the stratum of translation surfaces with one singular point of multiplicity 2 the Masur-Smillie constant is $1/2$.

Hausdorff dimension estimates in parameter space Instead of fixing a translation surface S and studying the set $\mathcal{N}\mathcal{U}(S)$, one can also study the set $\mathcal{N}\mathcal{U}(d, \pi)$ IETs of $d \geq 2$ intervals rearranged by π which are non-uniquely ergodic. Notice that for $d = 2, 3$ (since all irrational rotations are uniquely ergodic by Kronecker theorem) this set is countable and coincides with the set of non-minimal IETs.

Athreya and Chaika (2015) considered 4-IETs (with irreducible combinatorics π) and proved that the subset $\mathcal{N}\mathcal{U}(4)$ of 4-IETs which are not uniquely ergodic has Hausdorff dimension $5/2$. The construction of a rich subset of non-uniquely ergodic IETs in Δ_4 is a far-reaching generalization of Keane's non-uniquely ergodic 4-IET original example. This result was generalized by Chaika and Masur (2020) to any $d \geq 4$: they showed that if π belongs to the Rauzy class of the hyperelliptic permutation, then $\mathcal{N}\mathcal{U}(d, \pi)$ has Hausdorff dimension $d - 3/2$. Since the simplex Δ_d in (3) has dimension $d - 1$, this shows that the Hausdorff codimension of $\mathcal{N}\mathcal{U}(d, \pi)$ is exactly $1/2$ for any $d \geq 4$ (a result to be compared with the Hausdorff dimension $1/2$ of the set $\mathcal{N}\mathcal{U}(S)$ of non-uniquely ergodic directions on a given S , which in S^1 has

also Hausdorff codimension $1/2$). The result about IETs is used in Chaika and Masur (2020) to prove that the Masur-Smillie constant is $1/2$ for hyperelliptic components in the moduli space.

Behavior of Ergodic Averages

Given an ergodic area-preserving flow $\varphi_{\mathbb{R}}$ (or its restriction to an ergodic minimal component $S' \subset S$) and a real-valued observable f , let $I_T(f, x)$ denote the ergodic integral

$$I_T(f, x) := \int_0^T f(\varphi_t(x)) dt. \quad (8)$$

By (unique) ergodicity, if f has zero-mean, $\frac{1}{T}I_T(f, x)$ decay to zero for a.e. $x \in S$ (and all x with an infinite trajectory if f is continuous and f is zero in a neighborhood of the fixed points). If S is a torus, for a.e. linear flow (as well as a.e. minimal Hamiltonian flow) $I_T(f, x)$ is bounded and f is a *coboundary* (see below). Two distinctive features of these ergodic integrals for higher genus surfaces, i.e., if $g \geq 2$, is that there are *obstructions* for $I_T(f, x)$ to be a coboundary and that the *oscillations* are of *polynomial* nature.

Obstructions and cohomological equations - Understanding when the integrals $I_T(f, x)$ are bounded can be translated in a problem of solutions of cohomological equations.

The cohomological equation for smooth flows Let X be infinitesimal generator of the smooth flow $\varphi_{\mathbb{R}}$ preserving the smooth area μ . Let us recall that a mean zero observable $f: S \rightarrow \mathbb{R}$ is said to be a *coboundary* for $\varphi_{\mathbb{R}}$ if one can find a solution $u: S \rightarrow \mathbb{R}$ of the *cohomological equation* $Xu = f$. When u is bounded, this forces the ergodic integrals $I_T(f, x)$ to be bounded. When $g = 1$ and f is smooth, for a.e. linear flow f is a coboundary. In a seminal paper Forni (1997), Forni discovered the existence, for $g \geq 2$, of finitely many distributional obstructions for a regular f (in a suitable Sobolev space) to be a coboundary, given by g *invariant distributions*,

i.e., linear functionals $\mathcal{D}_i$, $1 \leq i \leq g$ on Sobolev observables. By *obstruction* one means here that if $\mathcal{D}_i(f) \neq 0$ for some $1 \leq i \leq g$, then f cannot be a coboundary. The value of the first distribution is simply the integral of $\int_S f d\mu$, so in genus $g = 1$ this is the only condition and is automatically satisfied for mean-zero observables. Furthermore, in Forni (2021) he shows that for a.e. smooth minimal area-preserving flow when f belongs to the kernel of all $\mathcal{D}_i$, $1 \leq i \leq g$, f is indeed a coboundary.

The cohomological equation for IETs Inspired by the work of Forni (1997) and Marmi et al. (2005) considered the cohomological equation over an IET T , rediscovered the obstructions in this setting, and described a full measure condition (that they called *Roth-type* condition) that guarantee that after removing the obstructions, the cohomological equation can be solved. More precisely, they show that for a Roth-type IET $T: I \rightarrow I$, given an observable $v: I \rightarrow \mathbb{R}$, absolutely continuous on each continuity interval I_i and with the derivative of bounded variation and $\int_I v'(x) dx = 0$ (zero sum of jumps), one can find $\chi: I \rightarrow \mathbb{R}$ piecewise constant on each I_i and a continuous $u: I \rightarrow \mathbb{R}$ such that

$$u \circ T - u = v - \chi. \quad (9)$$

The function χ , which belongs to the d -dimensional space, provides the obstruction for v to be a coboundary. In Marmi and Yoccoz's (2016), it was later shown that u is Hölder continuous. Further generalizations of this result were proved by Marmi et al. (2020) (who show how to modify the proof to assume a weaker condition, called *absolute Roth-type*); Forni et al. (2017) (who deal with the case of some zero Lyapunov exponents); and Lanneau et al. (2021) (who considered the case of T linear involution).

Power deviations of ergodic averages Zorich in the 1990s discovered experimentally (by considering IETs and Birkhoff sums of characteristic functions of continuity intervals) that for almost every linear flow in higher genus ($g \geq 2$), the integrals $I_T(f, x)$ (in the special case of observables corresponding to cohomology classes)

display *polynomial* oscillations, i.e., for almost every initial point $|I_T(f, x)| \sim O(T^\nu)$ in the sense that

$$\limsup_{T \rightarrow \infty} \frac{\log |I_T(f, x)|}{\log T} = \nu \quad (10)$$

for some exponent $0 < \nu < 1$. This phenomenon, known as *polynomial deviations* of ergodic averages, was soon after explained in seminal work by Kontsevich (1997) and Zorich (1997) relating power deviations to Lyapunov exponents of renormalization and proposed a finer description of polynomial deviations which became known as *Kontsevich-Zorich conjecture*.

The Kontsevich-Zorich conjecture Kontsevich & Zorich conjectured in Kontsevich and Zorich (1997) that the phenomenon of polynomial deviations holds for the ergodic integrals $I_T(f, x)$ of any smooth observable (and not only those coming from cohomology classes, which can be reduced to the study of Birkhoff sums over the IET obtained as Poincaré section of the flow for an observable that is a characteristic function of a continuity interval, i.e., the setting originally considered in Zorich (1997)). They conjectured furthermore that ergodic integrals display a *power spectrum* of oscillatory behaviors, i.e., there are exactly g positive exponents $0 < \nu_g \leq \dots \leq \nu_2 < \nu_1 := 1$ (which correspond to the positive Lyapunov exponents of renormalization) and, for each, a subspace of finite codimension of smooth observables that present polynomial deviations as above with exponent $\nu = \nu_i$. Forni (2002) could prove the bulk of this conjecture for linear flows on translation surfaces and for integrals of sufficiently regular functions, by showing there are indeed g positive exponents. His results also apply to locally Hamiltonian flows in $\mathcal{U}_{\min}$ and to sufficiently regular observables which vanish at the singularities. The *simplicity* of the spectrum, namely that the g exponents are all *distinct*, was later proved in Avila and Viana (2007).

Bufetov functionals and limit shapes A finer analysis of the behavior of Birkhoff sums or integrals, beyond the *size* of oscillations, appears in

Bufetov (2014). Bufetov (2014) shows in particular that (for *typical* translation flows and sufficiently regular observables) the *asymptotic behavior* of ergodic integrals can be described in terms of g (where g is the genus of the surface) *cocycles* $\Phi_i(t, x)$, $1 \leq i \leq g$ (also called *Bufetov functionals*): each $\Phi_i : \mathbb{R} \times S \rightarrow \mathbb{R}$ is a cocycle over the flow $\varphi_{\mathbb{R}}$ (in the sense that $\Phi_i(t + s, x) = \Phi_i(t, x) + \Phi_i(s, \varphi_t(x))$ for any $x \in S$ and $t \in \mathbb{R}$), $\Phi_1(T, x) \equiv T$ and each Φ_i has power deviations $|\Phi_i(T, x)| \sim O(T^{v_i})$ with exponent v_i . Together, the cocycles encode the *asymptotic behavior* of the ergodic integrals up to sub-polynomial behavior, in the sense that for some constants $c_i = c_i(f)$,

$$\begin{aligned} \int_0^T f(\varphi_t(x)) dt &= c_1 T + c_2 \Phi_2(T, x) + \dots \\ &\quad + c_g \Phi_g(T, x) + \text{Err}(f, T, x), \end{aligned} \quad (11)$$

where for every $x \in S$ with regular orbit, the quantity $\text{Err}(f, T, x)$ is an *error term* which grows subpolynomially, i.e., for any $\epsilon > 0$ there exists $C_\epsilon > 0$ such that

$$|\text{Err}(f, T, x)| \leq C_\epsilon T^\epsilon.$$

Using these results, Bufetov could also prove some limit theorems for translation flows, in particular the convergence along (exponentially sparse) subsequences of the distribution of ergodic integrals of regular observables, when the initial point $x \in S$ is randomized (see Bufetov (2014) for statements).

Deviations phenomena due to singularities Frączek and Ulcigrai (2021) gave new proof of the existence of a power deviation spectrum and asymptotic cocycles for smooth observables over locally Hamiltonian flows with Morse singularities which extends Bufetov-Forni results to smooth observables which do not vanish at singularities as well as flows in $\mathcal{U}_{\neg \min}$. Their approach, inspired by Marmi et al. (2005), provides also a description of the full measure set of locally Hamiltonian flows in terms of a Diophantine-like condition. Frączek and Kim (2021) pushed a similar approach to treat

the case of degenerate saddles and discovered the existence of a new type of power deviations associated to smooth observables that do not vanish at some degenerate saddles. The new exponents are determined by the jet of the observable at the degenerate singularity.

Mixing Properties

Finer chaotic properties such as mixing or weak mixing (as well as, more generally, spectral properties), crucially depend not only on the orbits of the flow, but also on the *speed* of motion along the orbits (i.e., the time-parametrization). In particular, they are very different for linear flows and their time-changes and locally Hamiltonian flows. For the latter, mixing or its absence depends crucially on the type of singularities.

Weak mixing in linear flows and IETs Weak mixing is an important and much investigated question for linear flows and IETs, in view of the lack of mixing in this setting.

Absence of mixing Katok proved already in the 1980s Katok (1980) that interval exchange transformations are *never* mixing (generalizing earlier work with Stepin; see Katok and Stepin (1967), [Remark 8.1] in the special case of 3-IETs). In the proof, Katok shows (using a simple combinatorial argument which exploits that the induced map of any d -IET on a subinterval is again an IET of at most $d + 2$ -subintervals) that any d -IETs is *partially rigid*, i.e., there exists $0 < \alpha < 1$ (depending on d only) and a diverging sequence of times $(n_k)_k$ such that for every measurable set $A \subset I$,

$$\text{Leb}(A \cap T^{n_k} A) \geq \alpha \text{Leb}(A), \quad \forall k \in \mathbb{N}. \quad (12)$$

From this it follows also that IETs have no mixing factor. In Katok (1980) it is furthermore shown, exploiting partial rigidity of IETs, that any special flow over an IET under a roof of bounded variation is not mixing. This implies in particular that linear flow, nor any smooth time-change of a linear flow, can be mixing. In the special case of special flows over rotations under smooth roofs

(or equivalently smooth time-changes of flows on $\mathbb{T}^2$) absence of mixing was proved earlier, in Kočergin (1972).

Full measure of weak mixing Recall that the ergodic theoretic property of weak mixing, by a classical result of Halmos (1960), is topologically generic while mixing is not. In Veech (1984) it was conjectured that almost every IET should be *weak mixing*, unless it is essentially a rotation (since rotations have eigenvalues and a discrete spectrum), i.e., if the combinatorial datum π is not that of a rotation with fake singularities. These IETs are called of *rotation type*.

Veech (1984) introduced a condition on the combinatorial datum π of an IET (called *W-condition*) under which he could prove full measure of weak mixing. As an example, when d is odd, the π which exchanges the order, i.e., maps $i \mapsto d - i + 1$ for $1 \leq i \leq d$ is of type W. Geometrically, the type W condition is essentially equivalent to asking that the IET is a Poincaré section of a flow on a (minimal component of a) surface which has *more than one* saddle point. Boshernitzan and Nogueira (2004), [Theorem 5.3] showed that all linearly recurrent IETs (or equivalently, bounded-type IETs, which have measure zero) with combinatorial datum of type W are weakly mixing. Notice that when π is not of type W, bounded-type IETs do not have to be weak mixing; see, for example, the examples of 4-IETs produced in Hmili (2010): these examples have π which is not type W and are not weak mixing, since they have a nonconstant eigenfunction. One can verify furthermore that they are linearly recurrent when the eigenvalue is badly approximable.

Nogueira and Rudolph could prove already in the 1990s that almost every IET with π not of rotation type is *topologically* weak mixing, i.e., has no nonconstant *continuous* eigenfunction (see Nogueira and Rudolph 1997) (in contrast with weak mixing, which is equivalent to the absence of nonconstant *measurable* eigenfunctions). An unrestricted proof of almost everywhere weak mixing for IETs not of rotation type, both for almost every linear flow and for almost every

IET, was given 10 years later by Avila and Forni (2007). It exploits deep results from Teichmüller dynamics, in particular positivity of the second Lyapunov exponent of the Teichmüller flow (Forni 2002) and a parameter exclusion argument, which is very delicate for the case of IETs, but much simpler in the case of linear flows (see the Appendix of Avila and Forni (2007)) on a.e. translation surface.

Weak mixing explicit examples The proof by Avila and Forni (2007) is not constructive. Explicit examples of weak mixing IET (of periodic type, i.e., self similar) were constructed in Sinai and Ulcigrai (2005) and Ferenczi and Zamboni (2011). Linearly recurrent (or bounded-type) IETs are also explicit (and include in particular periodic-type IETs); therefore also linearly recurrent IETs with combinatorial datum π of type W provide explicit examples of weak mixing IETs by Boshernitzan and Nogueira (2004).

Translation surfaces obtained by unfolding of a rational billiard (which constitute a measure zero subset) on which the linear flow is weak mixing for a.e. direction were constructed by Avila and Delecroix (2016) (in the class of Veech surfaces).

Hausdorff dimension of non-weak mixing Once it has been shown that weak mixing IETs or linear flows have measure zero, one can study the Hausdorff dimension of the complement. The parameter exclusion argument in Avila and Forni (2007) shows some estimates on the Hausdorff dimension of translation surfaces for which the vertical linear flow flows fail to be weak mixing. Avila and Leguil (2018) studied the Hausdorff dimension of the set of non-weakly mixing IETs. They showed that its Hausdorff dimension is strictly less than $d - 1$, where d is the number of exchanged intervals. It follows from Chaika and Masur (2020) that the Hausdorff dimension is at least $d - \frac{3}{2}$. Combining the work of Chaika and Masur (2020) and Al-Saqban et al. (2021), one can show that the set of non-weakly mixing IETs with permutation of type W has dimension $d - \frac{3}{2}$ (see Chaika and Masur (2020)).

Mild mixing in IETs and von Neumann flows Mild mixing, which can be defined as the absence of rigid factors, is an intermediate property between weak mixing and mixing. Thus, it is a natural property to explore systems which are known to be weak mixing but not mixing.

Mild mixing IETs As shown by Veech (1984), a typical IET is rigid, so it cannot be mildly mixing. The existence of mild mixing IETs has been first shown among 3-IETs: Ferenczi et al. (2005), [Theorem 4.1] showed that linearly recurrent 3-IETs are mild mixing (since they have a property known as *minimal self-joinings*; see Kanigowski and Lemańczyk (2020)). Robertson (2019) has considered IETs with combinatorics π of type W (the condition under which Veech first proved weak mixing) and shown that when linearly recurrent (a measure zero but full Hausdorff dimension condition), they are mild mixing.

Von Neumann flows Von Neumann flows in genus one (i.e., special flows over rotations or IETs under a piecewise absolutely continuous function with non-zero sum of jumps) were introduced by von Neumann (with rotations as the base) in 1936 to produce examples of weak mixing systems. He proved that such flows are weakly mixing for each irrational rotation if the roof function is piecewise C^1 . The weak mixing property was then proved for the von Neumann class of functions but over ergodic interval exchange transformations by Katok (2001). Frączek and Lemańczyk (2006) showed that von Neumann flows over rotations (with a roof with non-zero sum of jumps) can be furthermore *mild mixing* (but for rotation numbers α of bounded type, which form a measure zero set). This result was extended in Frączek et al. (2007) to roof functions with zero sum of jumps. Both results exploit versions of the Ratner properties (see below).

Mildly mixing linear flows As for IETs, the Veech result (Veech 1984) says that the linear flow on a typical translation surface is rigid, so it also cannot be mildly mixing. Frączek et al. (2007) gave a criterion for a piecewise constant

roof function over an irrational rotation with bounded partial quotients to be mild mixing. Using this criterion, Frączek (2009) has shown that when the genus is at least two, the set of Abelian differentials for which the vertical flow is mildly mixing is dense in every stratum of moduli space. He used an approximation of any translation surface by surfaces having many vertical saddle connections. Vertical flows on such surfaces have special representations over rotations.

Mixing properties of smooth flows In smooth area-preserving flows, due to the presence of Hamiltonian saddles, the closer a trajectory is to a saddle singularity, the more motion along the trajectory is slowed down. This generates a shearing phenomenon that can create mixing. The presence of absence of mixing depends both on the strength of the singularity (in particular if it degenerates or nondegenerates) and on the presence of traps (or more generally saddle loops homologous to zero), i.e., the *symmetry* or *asymmetry* of the roof in case of logarithmic singularities. For generic flows given by Morse forms one has indeed the following dichotomy:

Theorem 5 *Inside the open set $\mathcal{U}_{\min}$ in which the typical flow is minimal, almost every locally Hamiltonian flow is weakly mixing, but it is not mixing. On the other hand, for a full measure set of flows in $\mathcal{U}_{-\min}$, the restriction to each of minimal components is mixing.*

This statement summarizes a number of results (in particular Ulcigrai (2009, 2011) and Ravotti (2017)) which we now describe in detail.

Diophantine-type conditions Results on mixing properties require the introduction of Diophantine-like conditions, which describe the full measure set of locally Hamiltonian flows for which the results hold. For $g = 1$ these can be expressed as properties of the entries of the continued fraction expansion of the rotation number. In higher genus, these are expressed in terms of the renormalization given by the Teichmüller flow or renormalization algorithms for IETs such as Rauzy-Veech induction. For a survey of these

Diophantine-like conditions, we refer to the ICM proceedings by Ulcigrai (2022).

Mixing of Kocergin flows and related results The first examples of mixing smooth flows on higher genus surfaces were given already in the 1970s (in Kočergin (1975)) exploiting *degenerate saddles* (i.e., multi-saddles with $k \geq 6$ prongs, as in Fig. 4c) since in this case the saddles have a much stronger slowing down effect. Kochergin showed that special flows over ergodic IETs under roofs with polynomial singularities (see (7)) are mixing for all ergodic IETs on the base (in particular for all irrational rotation α in the base in the $g = 1$ case). This implies that a typical locally Hamiltonian flow is *mixing* if it has *degenerate saddles* (as in Fig. 4c) which all give rise to polynomial singularities of the roof function (e.g., for local Hamiltonians given by $\mathcal{J}(z^k)$ for $k \geq 2$).

As a special case, Kočergin (1975) result applies to flows on $\mathbb{T}^2$ with a stopping point (which can be seen as a fake degenerate saddle) that admit a representation as special flows over rotations under a roof with a symmetric power singularity. For this special case, polynomial mixing estimates were proven in Fayad (2001) for special observables when the power singularity is sufficiently strong.

While smooth time-changes of linear flows on $\mathbb{T}^2$ are never mixing by Katok (1980), a mechanism similar to the one exploited by Kočergin (1975) was used by Fayad (2002) to show that there exist smooth (actually analytic) time-changes of a flow on $\mathbb{T}^3$ which are mixing. Time-changes of flows on $\mathbb{T}^3$ can be seen as special flows over a two-dimensional rotation $\mathbb{R}_{\underline{z}} : \mathbb{T}^2 \rightarrow \mathbb{T}^2$, given by $(x_1, x_2) \mapsto (x_1 + \alpha_1, x_2 + \alpha_2) \in \mathbb{T}^2$. For mixing smooth time-changes to exist, $\underline{\alpha}$ should be a Liouville vector and the frequencies α_1, α_2 should have sequences of continued fraction denominators that are intercalated with respect to each other. Notice that these are measure zero examples.

Arnold conjecture and mixing in the asymmetric case Since the presence of a degenerate saddle is not generic, Kocergin work

motivated the question of mixing in locally Hamiltonian flows with only nondegenerate saddles (i.e., *simple* saddles). Arnold in the 1990s noticed a geometric phenomenon which could produce mixing in the ergodic minimal components of locally Hamiltonian flows on the torus known as Arnold flows (Arnold 1991). His intuition was proved to be correct shortly after by Sinai and Khanin (1992), who showed mixing under an arithmetic condition of full measure on the rotation number α . The full measure assumption on the rotation number α is that the entries a_n of the continued fraction expansion of α do not grow too fast, namely there exists a power $1 < \tau < 2$ and $C > 0$ such that $|a_n| \leq Cn^\tau$. The condition was later improved in Kochergin (2004).

The question of whether mixing is typical also for flows on higher genus is more delicate and requires a generalization of the Diophantine-like condition in higher genus, which was introduced in Ulcigrai (2007) and proved to be of full measure by Avila et al. (2006). Ulcigrai proved (Ulcigrai 2007) that for a.e. IET, the special flow under an asymmetric roof with one logarithmic singularities is mixing. Later, Ravotti (2017) showed that the same full measure condition can also be used to extend the result to roofs with several asymmetric logarithmic singularities, as in (4). This implies (see Ravotti (2017)) that in the open set $\mathcal{U}_{-\min}$, the restriction of the typical locally Hamiltonian flow $\varphi_{\mathbb{R}}$ on each of its minimal components is mixing.

In Ravotti (2017), *quantitative* results on the *speed* of mixing are also proved: for a typical $\varphi_{\mathbb{R}}$ in $\mathcal{U}_{-\min}$, restricted to a minimal component, the speed of decay of correlations (also sometimes called *speed of mixing*) is *sub-polynomial* and actually logarithmic, namely for every pair f, g of smooth observables there exist constants $c > 0, \alpha > 0$ such that

$$\left| \int (f \circ \varphi_t) \cdot g d\mu - \int f d\mu \cdot \int g d\mu \right| \leq c \log t^\alpha.$$

The role of shearing A crucial ingredient in the proofs of these mixing results (and more generally of mixing in parabolic flows) is played by a

geometric shearing phenomenon: if we consider a small arc γ transversal to the trajectories of the flow $\varphi_{\mathbb{R}}$, so that when flowing it, $\varphi_t(\gamma)$ passes nearby a saddle separatrix without hitting the saddle point (as shown in Fig. 13), the different deceleration rates of points cause $\varphi_t(\gamma)$ to *shear* in the direction of the flow (see Fig. 13a). Shearing allows to deduce mixing from ergodicity: by a Fubini-type argument, given measurable set $A \subset X$, for every large t one can cover an arbitrarily large proportion $A_t \subset A$ of with a collection of short transversal segments γ , each of which, after time t , shadows a long trajectory of $\varphi_{\mathbb{R}}$ and, therefore, by (unique) ergodicity is (close to) equidistributed. Furthermore, the speed of mixing depends on the speed of shearing.

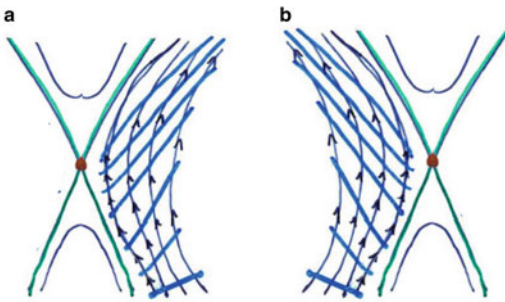
The shearing accumulated can be later destroyed when the segment $\varphi_t(\gamma)$ passes near the other side of a saddle (see Fig. 13b). The presence of a saddle loop, though, (as in Fig. 14a) typically creates an asymmetry (this was the key intuition of Arnold that had motivated his conjecture on mixing) by producing stronger

shearing on one side and hence, in this case, the accumulation of shearing predominantly in one direction produces global shearing.

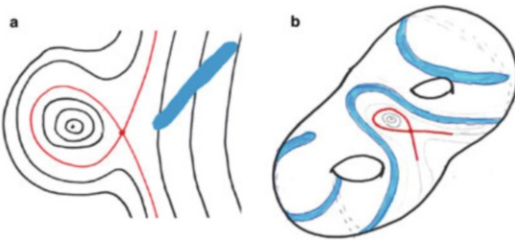
Ratner's properties Striking consequences of shearing (such as measure and joining rigidity) were proved for another famous class of parabolic flows, namely horocycle flows on hyperbolic surfaces and their time-changes, by exploiting a *quantitative shearing* property introduced by Marina Ratner and nowadays known as *Ratner property* (or RP). The difficulty in proving Ratner-type properties for smooth flows on higher genus surfaces is given by the presence of singularities (which are unavoidable when $g \geq 2$), which introduce discontinuities and destroy the slow form of divergence à la Ratner: as soon as two nearby trajectories are separated by hitting a saddle, indeed, one drastically loses control of the divergence. A first generalization of the Ratner property (the finite Ratner property) was introduced by Frączek and Lemańczyk (2006) and proved for a special class of von Neumann flows (special under a piecewise absolutely continuous roof with non-zero sum of jumps over an irrational rotation of bounded type). The finite Ratner property was also proved for flows when the sum of jumps is zero by Frączek et al. (2007). Notice that von Neumann flows are not (globally) smooth. For other examples of this type of Ratner property, see also Frączek and Lemańczyk (2010). A version of the Ratner property is also a key ingredient in the proof in Kanigowski (2015).

The Ratner property in its classical form as well as the weaker versions defined in Frączek and Lemańczyk (2006, 2010) is expected to fail for smooth area-preserving flows with non-degenerate fixed points. The failure of the classical Ratner property was formally proved in a special case in Fayad and Kanigowski (2016), [Theorem 1] (for a class of Kochergin flows), and this result gives reasons to believe that, for similar reasons, the classical Ratner property should indeed always fail in presence of singularities.

A new variant of the RP which has the same dynamical consequences, called *Switchable Ratner Property* (or SRP for short), was



Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 13 Shearing mechanism



Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 14 Mixing mechanism

introduced by Fayad and Kanigowski (2016) and proved to hold for almost every (minimal component of) an Arnold flow in $\mathcal{U}_{-\min}$ when $g = 1$ (as well as for a measure zero set of Kochergin flows). According to this variation, it is sufficient to see the Ratner divergence of orbits for most pairs of initial conditions (x, y) either in the future (for $t > 0$) or in the past (for $t < 0$), depending on the pair of initial points. Thus, if one pair of nearby trajectories is separated by hitting a singularity, and hence their distance explodes in an uncontrolled manner, one can still hope to *switch the direction of time* (from which the name of *switchable* Ratner property) and be able to prove the Ratner slow form of divergence when flowing backward in time. Kanigowski et al. (2019) generalize the result by Fayad and Kanigowski to higher genus and show that for almost every $\varphi_{\mathbb{R}} \in \mathcal{U}_{-\min}$, the restriction of $\varphi_{\mathbb{R}}$ to any minimal component satisfies the Switchable Ratner Property for any genus $g \geq 1$. This also covers as special case minimal components of typical flows in $g = 1$ with more than one simple saddle. In Fayad and Kanigowski (2016) the case of more saddles is also considered, but under a condition on the relative position of the saddles which has measure zero.

Multiple mixing Applying joining techniques introduced by Host (1991) and developed by Ryzhikov and Tuveno (2006), the Switchable Ratner property can be used in particular to show that mixing can be upgraded to *mixing of all orders*. In particular, the results in Fayad and Kanigowski (2016), combined with Sinai and Khanin (1992), imply that typical Arnold flows are *mixing of all orders* (as well as a measure zero set of Kochergin flows in genus one). Similarly, the result of Kanigowski et al. (2019) combined with Ravotti (2017) shows that for a full measure set of locally Hamiltonian flows in $\mathcal{U}_{-\min}$, each restriction to a minimal component is mixing of all orders.

Absence of mixing in the symmetric case For minimal locally Hamiltonian flows in $\mathcal{U}_{\min}$, which have only simple saddles, since there are no saddle loops homologous to zero and hence no

asymmetry which produces global shearing, one can expect that the effect of shearing on two different sides of the same saddles compensates and cancels out.

Symmetric logarithms over rotations Consider a special flow over an irrational rotation R_a under a roof with a symmetric logarithmic singularity, i.e.,

$$r(x) = C | \log(x) | + C | \log(1-x) |. \quad (13)$$

Already in the 1970s, Kočergin (1976) proved that these flows are *not* mixing for a full measure set of rotation numbers a and, much more recently, extended this result to *all* irrational rotation numbers (Kochergin 2007). The criterion for the absence of mixing exploited in Kočergin (1976) can be seen as a generalization of Kočergin (1972) absence of mixing result and shows that (at least for typical) locally Hamiltonian flows mixing via shearing is essentially the *only* possible way of achieving mixing. Indeed it was conjectured in Lemańczyk (2000) and proved by Schmidt (2002) that mixing in special flows over a rotation is only possible if the distribution of the sequence of Birkhoff sums of the roof function is not tight, i.e., $\text{Leb}(\{x; |r^{(n)}(x)| > K\})$ tends to 1 as K tends to infinity.

Symmetric logarithms over IETs While the key example studied by Kochergin (special flows over rotations under a symmetric logarithmic singularity) constitutes a prototype result for the absence of mixing, these special flows do not arise as representations of typical locally Hamiltonian flows, since special representations of typical minimal flows with only simple saddles always yield special flows under roofs with symmetric logarithmic singularities over IETs with $d \geq 4$. Ulcigrai (2011) proved the absence of mixing for such flows for almost all interval exchange transformations in the base. Thus, a.e. flow in $\mathcal{U}_{\min}$ is not mixing. A special case of the absence of mixing result for surfaces with $g = 2$ and two isometric saddles was proved in Scheglov (2009); see also the recent work by Chaika et al. (2021) for a more geometric proof. The proof in Scheglov (2009) exploits a combinatorial analysis

of the substitutions induced by Rauzy-Veech induction to show that orbits satisfy a certain symmetry. In Chaika et al. (2021), this symmetry is deduced geometrically from the existence of a hyperelliptic involution.

We remark that in $\mathcal{U}_{\min}$ there exist nevertheless exceptional mixing flows (contrary to the case of the torus in view of Kočergin (2007)), as shown in Chaika and Wright (2019), where the authors could produce sporadic mixing examples in $g = 5$. Also for this example, shearing is still at the base of mixing, but it is not produced by asymmetry of the singularities, but rather by an asymmetric equidistribution, so that trajectories, at different time scales, spend much more time on one side of a saddle than another. The IETs considered are finite covers of rotations (for which Diophantine-type conditions can be given in terms of continued fraction entries) which are only barely uniquely ergodic and for which orbits equidistribute very slowly and in a very asymmetric way.

More recently, Kanigowski and Kułaga-Przymus (2016) showed (exploiting the former work of Kułaga (2012)) that special flows with symmetric logarithmic singularities over IETs of bounded type are mild mixing (the assumption on the IET used in Kanigowski and Kułaga-Przymus (2016) is not explicitly that the IET is of bounded-type, but a condition on orbits of discontinuities, which can a posteriori be shown to be equivalent to bounded type or linearly recurrent). The main component of the proof is to show that the flow satisfies the Switchable Ratner Property.

Weak mixing via logarithmic singularities

Frączek and Lemańczyk (2003) showed that flows over all irrational rotations under any roof functions with one symmetric logarithm are weakly mixing. Generalizing this result, Ulcigrai (2009) also proved weak mixing for special flows with logarithmic singularities (not necessarily symmetric, although in the asymmetric case mixing is already known; see Ulcigrai (2007) and Ravotti (2017)) for almost every IET on the base. The proof exploits the partial rigidity of IETs proved by Katok (1980) as well as the presence of

logarithmic singularities. Thus, a.e. flow in $\mathcal{U}_{\min}$ is weakly mixing but not mixing.

Spectral Properties

We move now to the properties of the spectrum of Koopman unitary operators and their spectral measures. We recall that the Koopman unitary operator associated to a dynamical system is given by $f \mapsto f \circ T$ for $f \in L^2(I, dx)$ when T is an IET or $f \mapsto f \circ \varphi_t$ for $f \in L^2(S, \mu)$ when $\varphi_{\mathbb{R}}$ is a surface flow preserving μ . We refer the reader to the entry by Kanigowski and Lemańczyk (2020) or Katok and Thouvenot (2006) for a survey of spectral theory of dynamical systems.

Spectral measures of IETs and linear flows

The first general spectral result about IETs is Oseledec's theorem in Oseledec (1966) stating that the maximal spectral multiplicity of any d -interval exchange transformation T is bounded by d . In 1984, Veech proved in Veech (1984), [Theorem 1.4] the a.e. IET is rigid which implies by standard arguments (see, e.g., Katok and Thouvenot (2006)) that T has *singular spectrum*. It follows furthermore from Avila and Forni (2007) work on weak mixing that a.e. IET, as well as a.e. linear flow, has continuous spectrum. Exotic examples of IETs with eigenvalues with various properties were constructed by Ferenczi and Zamboni (2011)); see also Hmili (2010).

Let $\chi = \chi_A$ be the characteristic function of a continuity interval $A = I_i$ of T (or, more generally, of a floor of a cutting and stacking presentation of T , as those given by Rauzy-Veech induction). A condition for the spectral measure σ_χ to be continuous at a point $\omega \in S^1$ was given by Bufetov et al. (2006) and used to produce explicit examples of (periodic-type) IETs with continuous spectrum in Sinai and Ulcigrai (2005).

Sinai raised the question: to find modulus of continuity for the spectral measures of translation flows. In Bufetov and Solomyak (2018, 2020), Bufetov and Solomyak developed an approach to this problem and succeeded in obtaining Hölder estimates for spectral measures in the case of

surfaces of genus 2. The final result for surfaces of higher genus, independently proved in Bufetov and Solomyak (2021) and Forni (2022), says that for every stratum in the moduli space there are $\gamma > 0$ and $\beta > 0$ such that for a.e. translation surface in the stratum there exists $C > 0$ such that for every $f: S \rightarrow \mathbb{R}$ smooth enough

$$\sigma_f([\lambda - r, \lambda + r]) \leq C \left(\frac{1 + |\lambda|}{|\lambda|} \right)^\beta \|f\| r^\gamma \quad (14)$$

for every $\lambda \neq 0$ and $r > 0$. A similar result for a.e. IETs is harder to prove; Avila, Forni, and Safaee estimated moduli of continuity of spectral measures for a.e. IETs in Avila et al. (2021b).

Quantitative weak mixing The main tool leading to the proof (14) is quantitative estimates of so-called twisted ergodic integrals of the form

$$\left| \frac{1}{T} \int_0^T e^{2\pi i \lambda t} f(\varphi_t x) dt \right| \leq C \left(\frac{1 + |\lambda|}{|\lambda|} \right)^\beta \|f\| T^{-\gamma}$$

proved (in different forms) in Forni (2022), Bufetov and Solomyak (2021), and Avila et al. (2021b). On the other hand, quantitative estimations of twisted integrals lead to an efficient version of weak mixing

$$\frac{1}{T} \int_0^T |f(\varphi_t x), g|^2 dt \leq C \|f\| \|g\| T^{-\gamma}$$

for f, g smooth enough zero mean functions; see Forni (2022) (and Avila et al. (2021b) for IETs).

Spectra of locally Hamiltonian flows While the classification of mixing properties of locally Hamiltonian flows is essentially complete (see Theorem 5), there are few results on their spectral properties for $g \geq 2$.

First examples with singular spectrum Frączek and Lemańczyk (2003) showed singularity of the spectrum for special flows under a symmetric logarithmic roof over a full measure set of rotations (this follows from *spectral disjointness* from all mixing flows). They also use this result to build examples of locally Hamiltonian flows on

surfaces of any genus $g \geq 2$ with singular continuous spectrum (see Frączek and Lemańczyk (2003), [Theorem 1]) using Blohin (1972) construction. Since these are essentially built gluing genus one flows and the resulting flow has a lot of saddle connections, they are highly non-typical.

Lebesgue spectrum via a stopping point A perhaps the most surprising result for minimal flows on the torus with one stopping point or *fake* singularity (see Fig. 15) was proved recently by Forni, Fayad, and Kanigowski. In Fayad et al. (2021), they proved that if the *degenerate* singularity is sufficiently strong, the spectrum is *countable Lebesgue*. In Kočergin (1975), Kocergin flows admit a special flow representation where the roof has power-type singularities. In genus one, for a flow with one degenerate singularity, one has a special flow over a rotation, under the roof $r(x) = c_0/x^\gamma + c_1/(1-x)^\gamma$ for some $0 < \gamma < 1$. The assumption in Fayad et al. (2021) is that γ is sufficiently close to 1. The absolute continuity of the spectrum, taking as a starting point an idea from Forni and Ulcigrai (2012) is proved from estimate on the (polynomial) speed of decay for natural observables, by showing that decay is square-summable. Forni and Ulcigrai exploited square-summable estimates on the decay of correlations to prove in Forni and Ulcigrai (2012) that smooth time-changes of horocycle flows have the Lebesgue spectrum, as conjectured by Katok and Thouvenot (2006). An extra difficulty in this setting is given by the symmetry of the power singularity, which creates cancellations and parts of space where there is no shearing. The authors



Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 15 $g = 1$, stopping point

also introduce a criterium to show that the spectrum has countable multiplicity suited for parabolic settings. The criterium was also used in Fayad et al. (2021) to complete the proof of the Katok-Thouvenot conjecture on smooth time-changes of horocycle flows, by showing countable multiplicity of the Lebesgue spectrum.

Singular spectrum in genus 2 In the opposite direction, Chaika et al. proved in Chaika et al. (2021) that a *typical* minimal locally Hamiltonian flow on a *genus two* surface with two isomorphic *simple saddles* (Fig. 16) has *purely singular* spectrum. The result, inspired by the techniques used to prove the singularity result (for special flows over rotations) in Frączek and Lemańczyk (2003, 2004), deduces singularity of the spectrum from absence of mixing and rigidity, exploiting the geometric symmetries given by the hyperelliptic involution. It also provides an independent proof of absence of mixing for typical flows in the same class ($g = 2$, two isomorphic saddles, see Fig. 16) proved in Scheglov (2009). As in Scheglov (2009), the assumption that the saddles are isometric is crucial to guarantee that the underlying surface has an inner symmetry, which plays a crucial role in the proof. More precisely, the linear flow of which the locally Hamiltonian flow is a time-change is a flow on a translation surface S which admits a *hyperelliptic involution*, i.e., an affine automorphism $\Phi : S \rightarrow S$ which is an involution, i.e., $\Phi^2 = Id$.

Disjointness Results

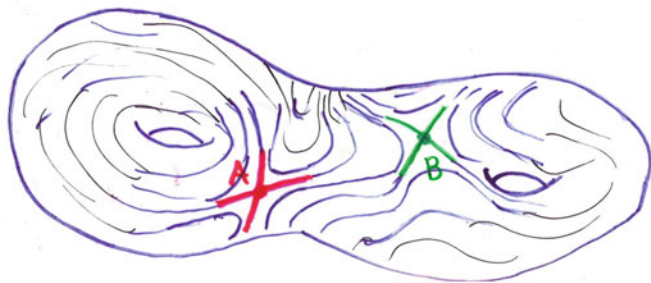
The notion of *disjointness* was introduced in the 1960s by Furstenberg (see in particular Furstenberg (1967)). Recall that disjoint flows

are not isomorphic and, moreover, their factors are also not isomorphic. We refer to the survey de la Rue (2020).

Disjointness from mixing systems For all systems for which absence of mixing is known, one can try to strengthen the result by proving disjointness from all mixing flows. The proof of that no IET is mixing in Katok (1980) already implies (even if this is not explicitly stated) that IETs are disjoint from all mixing systems; see also Ryzhikov (1994). In fact, partial rigidity (see (12)) implies disjointness from mixing systems. The proof by Katok also covers the case of special flows over IETs under piecewise constant roofs, which are also partially rigid. This strategy does not work for all roof functions of bounded variation. Frączek and Lemańczyk (2009a), [Appendix A] showed that von Neumann flows over IETs are never partially rigid.

Frączek and Lemańczyk strengthened Kočergin results in Kočergin (1972, 1976) showing absence of mixing for flows over rotations (for smooth roofs and symmetric logarithmic singularities respectively) by proving in Frączek and Lemańczyk (2004) that for every irrational α , the special flows over R_α under a smooth roof either has bounded variation or whose Fourier coefficients decay as $O(1/n)$ (so in particular for a roof with one symmetric logarithmic singularity) are disjoint from all mixing flows. Spectral disjointness was also proved in Frączek and Lemańczyk (2003) for a.e. rotation and the roof function with one symmetric logarithmic singularity. All these results are based on tightness of the distribution of Birkhoff sums of the roof function, and the study of their weak limits of joinings (via Markov operators) determined the graphs of

Ergodic and Spectral Theory of Area-Preserving Flows on Surfaces, Fig. 16 $g = 2$, isomorphic saddles



φ_t in $S \times S$. This approach was strengthened in Frączek and Lemańczyk (2005), where disjointness from all mixing systems was proved also for all special flows over ergodic IETs and under bounded variation roofs (in particular for all von Neumann flows over IETs) and for special flows over a.e. rotation and under any roof function with symmetric logarithmic singularities (in particular for Blokhin flows in any genus).

Notice that the tightness proved by Ulcigrai (2011) is sufficient to apply the techniques developed in Frączek and Lemańczyk (2005). Combining these two results one has that a.e. locally Hamiltonian flow in $\mathcal{U}_{\min}$ is disjoint from all mixing flows.

Joining properties of IETs As we have already mentioned, every IET is partially rigid, so is disjoint from all mixing systems. Moreover, a.e. IET is rigid; see Veech (1984). Chaika (2012) proved a much stronger result about rigidity of a.e. IET, saying that if $(q_n)_{n \in \mathbb{N}}$ is any increasing sequence of natural numbers, then a.e. IET is rigid along a subsequence of $(q_n)_{n \in \mathbb{N}}$. This implies an amazing result that any ergodic system is disjoint from a.e. IET; see Chaika (2012), [Theorem 1].

Veech asked, already in the 1980s (see Veech (1982a)), whether a.e. IET is *simple*, i.e., the only (ergodic) self-joinings are graphs and the product measure. This question was motivated by the article by del Junco (1983), where he found a one-parameter family of simple 3-IETs. Further examples of simple d-IETs with $d \geq 3$ were produced in Ferenczi and Zamboni (2011). A surprisingly negative response to Veech's question was given by Chaika and Eskin (2021) where they studied the set of joinings of a.a. 3-IETs and proved that it is rather large. The set of self-joinings is a Poulsen simplex, i.e., the set of ergodic joinings is dense.

Non-reversibility A measure-preserving flow $\varphi_{\mathbb{R}} = (\varphi_t)_{t \in \mathbb{R}}$ is *reversible* if it is isomorphic to its inverse flow $\varphi_{\mathbb{R}}^{-1} = (\varphi_{-t})_{t \in \mathbb{R}}$. It is also sometimes assumed that the map giving the isomorphism is an involution (its second iteration is the identity). Reversibility is often observed for dynamical systems of physical origin. In contrast, from the point of view of ergodic theory, this is a

rare property. As shown by del Junco (1981) (for automorphisms) and by Danilenko and Ryzhikov (2012) (for flows), typically we have even disjointness with the inverse action. Although non-reversibility is a typical property, it is not easy to prove in concrete examples, especially for flows on surfaces. In general, proving the non-isomorphism of zero entropy and spectrally equivalent systems (a flow is always spectrally isomorphic with its inverse) is a difficult task. One of the standard techniques here is disjointness, for which proving techniques are better developed.

The first result for surface flows was proved in Frączek and Lemańczyk (2009a), where, using a shearing mechanism and Ratner's techniques, the authors showed that for every rotation R_α of bounded type, every von Neumann flow is disjoint from its inverse. This result was expended to a.a. rotations in Frączek et al. (2014) and to a.e. IET in Berk and Frączek (2015), but in both papers only the non-isomorphism of the von Neumann flow and its inverse is proved. The methods developed in Frączek et al. (2014) and Berk and Frączek (2015) (do not rely on Ratner's techniques) are inspired by Ryzhikov's results and involve the study of weak limits of graph 3-joinings. They were further creatively developed in Berk et al. (2020), where they were allowed to prove disjointness of von Neumann flows with their inverse for a.a. IETs.

The problem of reversibility in a class of linear flows on translation surfaces is more complicated. Here the answer depends on connected component C in the moduli space. If the connected component C is hyperelliptic, then every linear flow on any translation surfaces in C is isomorphic to its inverse via the hyperelliptic involution. On the other hand, for a typical (in the topological sense) translation surface in any non-hyperelliptic component C , the linear flow is disjoint from its inverse; see Berk et al. (2020).

Disjointness of rescalings A property which seems to be common among parabolic flows (apart from some well-known exceptions in the homogeneous world) is *disjointness of rescalings*, defined as follows. Given a real number $\kappa > 0$, by

the κ -rescaling of $\varphi_{\mathbb{R}} = (\varphi_t)_{t \in \mathbb{R}}$, we simply mean the flow $\varphi_{\mathbb{R}}^{\kappa} := (\varphi_{\kappa t})_{t \in \mathbb{R}}$ (in which the time is rescaled by the factor κ). Thus, a rescaling is a special type of time-reparametrization of a flow, given by a linear time-change. We say that $\varphi_{\mathbb{R}}$ has *disjoint rescalings* if for all (or all but finitely many) $p, q > 0$, the rescalings $\varphi_{\mathbb{R}}^p$ and $\varphi_{\mathbb{R}}^q$, where $p, q > 0$ and $p \neq q$, are disjoint (in the sense of Furstenberg). Disjointness of rescalings has played a key role in proving some of the first instances of Sarnak's conjecture on the Moebius function (see the survey Kułaga-Przymus and Lemańczyk (2020)).

Rescalings of Arnold flows Kanigowski, Lemańczyk, and Ulcigrai recently proved in Kanigowski et al. (2020) that disjointness of rescalings is typical among Arnold flows (see Fig. 3b):

Theorem 6 *Let $\varphi_{\mathbb{R}}$ be the (Arnold) special flow over a.e. rotation R_{α} under a roof with asymmetric logarithmic singularities with constants $C_0 \neq C_1$ given by $r(x) = C_0 |\log x| + C_1 |\log(1-x)| + h(x)$ where h is smooth. Then there exist only two values of the form $q, 1/q$ such that $\varphi_{\mathbb{R}}$ and $\varphi_{\mathbb{R}}^p$ are disjoint for any positive $p \notin \{1, q, 1/q\}$, where $q = C_0/C_1$.*

The proof exploits a new criterium for disjointness based on the *switchable* Ratner property. The criterion was devised and formulated so that it can be applied to prove disjointness of two flows which both have the switchable Ratner property when in both one can observe a controlled form of divergence of nearby trajectories (e.g., polynomial divergence), but the speed of divergence for the two flows is different (e.g., for one flow it is linear, and in the other, quadratic).

Disjointness from other parabolic flows The criterion is used in Kanigowski et al. (2020) also to show that a typical Arnold flow is disjoint from any smooth time-change of the horocycle flow (and in particular from the classical horocycle flow itself), thus showing that these two classes of parabolic flows are truly distinct.

Rescalings of von Neumann flows Using simple joining arguments Frączek and Lemańczyk (2009a) showed that every von Neumann flow over any ergodic IET is not isomorphic to its any rescaling. Disjointness of rescalings for von Neumann flows was studied by Berk and Kanigowski (2021). They considered von Neumann flows over IETs and showed disjointness (even spectral disjointness) of rational rescalings, i.e., rescalings $\varphi_{\mathbb{R}}^{\kappa}$ where $\kappa \in \mathbb{Q}$, for a.e. IET. A similar proof also allows dealing with special flows over an IET T under piecewise constant function with a singularity in the interior of a continuity interval of the IET. Spectral disjointness of rational rescalings is also proved in Berk and Kanigowski (2021) for a.e. IET.

Von Neumann flows over rotations The new criterion for disjointness introduced by Kanigowski et al. (2020) has been also used by Dong and Kanigowski (2020) to study von Neumann flows in genus one. They proved that for any irrational rotation R_{α} , any real rescaling of a special flow $\varphi_{\mathbb{R}}$ over R_{α} under a piecewise absolutely continuous function with only one discontinuity is disjoint from $\varphi_{\mathbb{R}}$.

Symmetric log over rotations Berk and Kanigowski (2021) also studied the case of a roof with one symmetric logarithmic singularity over rotations (Kochergin prototype example of the absence of mixing) and also proved that in this case one has (spectral) disjointness of rational rescalings for a.e. rotation number. The techniques in this symmetric setting (where no form of Ratner property is known) are very different and based on a refinement of the techniques used to prove the absence of mixing and disjointness from mixing flows in Frączek and Lemańczyk (2004).

Open Directions

Despite the surge of activity and many results on ergodic, mixing, and spectral properties which have been proved in the last decades on area-preserving flows on surfaces, both on IETs and linear flows exploiting Teichmüller dynamics and on smooth or locally Hamiltonian flows, many

questions remain open. We indicate here only some questions and directions which arise naturally from the results surveyed in this entry.

Linear flows and IETs Even though the ergodic theory of typical IETs and translation flows is much studied and well understood, some questions remain open. Weak mixing is known for a.e. IET and a.e. linear flow but is not explicit and while explicit examples of weak mixing (and its lack) were constructed, a classification of all translation surfaces on which linear flows are weak mixing for a.e. direction is still out of reach. Another open question is the exact computation of Hausdorff dimension of the set of non-weakly mixing IETs (see Question 1.6 in Chaika and Masur (2020)).

Another Hausdorff dimension computation question concerns non-uniquely ergodic IETs, namely what is the Hausdorff dimension of d -IETs with precisely k ergodic invariant probability measures, for a given $1 < k \leq g \leq d/2$ (listed as Question 1.8 in Chaika and Masur (2020)).

While a.e. IET has singular spectrum, whether there can exist exceptional IETs with an absolutely continuous component in the spectrum is not known.

The simplicity conjecture in Veech (1982a) on simplicity of d -IETs for $d \geq 4$ is still wide open, together with an understanding of joinings and self-joinings. It also remains open whether almost every 3-IET is prime; it was a question suggested by Veech. A measure-preserving system is called prime if it has no non-trivial factors.

While upper bounds on the Hausdorff dimension of non-uniquely ergodic IETs are now known, a finer understanding of the set of non-uniquely ergodic directions $\mathcal{N}\mathcal{U}(S)$ on a given surface S , in the spirit of the full dichotomy for the slit-tori example, is probably out of reach. Even if there exists a surface for which the Hausdorff dimension of $\mathcal{N}\mathcal{U}(S)$ is strictly between 0 and $1/2$ is not known (see also Question 1.9 in Chaika and Masur (2020)).

Locally Hamiltonian Flows

Exceptional mixing results Nevertheless, many questions about exceptional behavior can be

further investigated. While the typical flow in $\mathcal{U}_{\min}$ is weakly mixing but not mixing (see Theorem 5), we do not know if weak mixing holds everywhere when $g \geq 1$ and no other examples other than the rather special examples of mixing flow in $g = 5$ in Chaika and Wright (2019) were constructed. Within $\mathcal{U}_{\min}$, mixing holds a.e. on every minimal component (see Theorem 5), but one can wonder if there exist exceptional examples without mixing for $g \geq 2$. They are known not to exist in $g = 1$ in view of Kochergin work Kochergin (2004), where mixing is proved for every irrational rotation and the roof function with logarithmic singularities of strongly asymmetric type.

As we discussed, a.e. flow in $\mathcal{U}_{\min}$ has not only minimal components which are mixing, but also mixing of all orders. Proving that *any* mixing minimal component of a flow in $\mathcal{U}_{\min}$, as well as any mixing Kochergin flow, is also mixing of all orders is an open problem (see Question 38 in Fayad and Krikorian (2018) ICM proceedings). Showing it would confirm the famous (still open) Rokhlin conjecture (i.e., that mixing implies mixing of all orders; see Quas (2017)) in the setting of area-preserving surface flows.

Finally, all examples where mild mixing is known (for special flows over rotations or IETs under symmetric logarithmic singularities, but also for von Neumann flows and for IETs themselves) are over IETs which are linearly recurrent (or bounded-type) rotations or IETs. Perhaps mild mixing is not possible otherwise.

Nature of the spectrum Spectral property of locally Hamiltonian flows is a natural question, which has been lingering for decades (see, e.g., Katok and Thouvenot (2006), [Section 5]).

The recent result by Fayad et al. (2021) for flows in $g = 1$ with stopping points suggests that it may be possible to prove that the spectrum is countable Lebesgue also in higher genus when in presence of degenerate, sufficiently strong (multi-saddle) singularities. It is not clear what to expect when the degenerate singularity is not sufficiently strong. In Fayad et al. (2021), the power γ of the singularity of the roof of the special flow representation (see (7)) is close to 1. One might hope

that absolute continuity could hold for all powers $\gamma \geq 1/2$, but this is out of reach with the current techniques. If $\gamma < 1/2$, the approach which uses square-integrability of the decay of correlations estimates fails completely.

In the nondegenerate case, the singularity result proved in Chaika et al. (2021) for typical flows in $g = 2$ with simple (isomorphic) *saddles* indicates that the spectrum could be *purely singular* also for a.e. flow in $\mathcal{U}_{\min}$. In the open set $\mathcal{U}_{\min}$, which consists of flows with nondegenerate singularities that are not minimal, but have *several* minimal components, the nature of the spectrum (for the restriction of a typical flow to a minimal component) is unclear. These flows are indeed mixing, but with sub-polynomial rate (see Ravotti (2017), which provides logarithmic upper bounds), and it is not clear whether to expect singularity or absolute continuity of the spectrum. The nature of the spectrum is an open problem even in $g = 1$, for Arnold flows.

Disjointness of rescalings We believe disjointness of rescalings should also hold for typical locally Hamiltonian flows in higher genus, but this is currently an open problem. Preliminary work seems to indicate that, despite some technical additional difficulties, the techniques used to prove Theorem 6 should allow proving disjointness of rescalings for all minimal components of typical flows in the open set $\mathcal{U}_{\min}$.

Berk and Kanigowski (2021) proof of disjointness of (rational) rescalings for *symmetric* logarithmic singularity over a full measure set of rotations (Kochergin prototype example of absence of mixing) gives a good indication that disjointness of rescalings could also hold for typical flows (under a suitable full measure Diophantine-type condition) in the complementary set $\mathcal{U}_{\min}$.

Non-compact surfaces We restricted in this survey to the discussion of the ergodic theory of flows on surfaces which are compact. Even though it is just at its beginning, the study of *infinite* translation surfaces (i.e., translation surfaces of infinite area or infinite genus) and their

linear flows, as well as infinite extensions of locally Hamiltonian flows, has been an active recent area of research. The latter extensions, as well as the study of translation surfaces which are Abelian covers of compact translation surfaces, is related to the study of their Poincaré maps, which are skew-product extensions of IETs, which provide important infinite ergodic theory examples and generalize the much studied theory of piecewise smooth cocycles over rotations. A skew-product extension of $T : I \rightarrow I$ over a group G given by the cocycle $\varphi : I \rightarrow G$ is the map $(x, g) \mapsto (Tx, g + \varphi(x))$. Such skew products where $G = \mathbb{Z}^d$ and φ is piecewise constant appear as Poincaré maps of Abelian covers of translation surfaces; $\mathbb{R}$ -extensions of locally Hamiltonian flows provide cocycles φ which are piecewise continuous or have (logarithmic or power-like) singularities.

Many results on the ergodicity of such infinite measure-preserving flows and maps were proved for special families of examples; see, e.g., Hubert and Weiss (2013), Hooper (2015), Málaga Sabogal and Troubetzkoy (2016), Ralston and Troubetzkoy (2017), or Chaika and Robertson (2019) for linear flows and IETs with infinite invariant measures and Fayad and Lemańczyk (2006), Conze and Frączek (2011), or Frączek and Ulcigrai (2012, 2021) for extensions of locally Hamiltonian flows. Some typical non-ergodicity results were proved as well; see Frączek and Ulcigrai (2014) and Frączek and Hubert (2018). The problem of ergodicity is nevertheless still widely open, both in the context of linear flows on Abelian covers of translation surfaces and for extensions of locally Hamiltonian flows. Other possible directions of investigation are the study of Radon invariant measures and the existence of limit theorems.

Acknowledgments K. F. acknowledges the support of the Narodowe Centrum Nauki Grant number 2017/27/B/ST1/00078. C. U. is part of SwissMAP (*The Mathematics of Physics* National Centre for Competence in Research) and is currently supported by a SNSF (Swiss National Science Foundation) Grant number 200021_188617/1. Both are acknowledged for their support.

Bibliography

- Al-Saqban H, Apisa P, Erchenko A, Khalil O, Mirzadeh S, Uyanik C (2021) Exceptional directions for the Teichmüller geodesic flow and Hausdorff dimension. *J Eur Math Soc (JEMS)* 23(5):1423–1476
- Aranson SH, Grines VZ (1973) Certain invariants of dynamical systems on two-dimensional manifolds (necessary and sufficient conditions for the topological equivalence of transitive systems). *Mat Sb (NS)* 90(132):372–402, 479,
- Arnold VI (1991) Topological and ergodic properties of closed 1-forms with incommensurable periods. *Funktsional'nyi Analiz i Ego Prilozheniya* 25(2): 1–12. (Translated in: *Funct Anal Appl* 25(2):81–90, 1991)
- Athreya JS, Chaika J (2015) The Hausdorff dimension of non-uniquely ergodic directions in $H(2)$ is almost everywhere $\frac{1}{2}$. *Geom Topol* 19(6):3537–3563
- Avila A, Delecroix V (2016) Weak mixing directions in non-arithmetic Veech surfaces. *J Am Math Soc* 29(4): 1167–1208
- Avila A, Forni G (2007) Weak mixing for interval exchange transformations and translation flows. *Ann Math* 165(2):637–664
- Avila A, Forni G, Ravotti D, Ulcigrai C (2021a) Mixing for smooth time-changes of general nilflows. *Adv Math* 385. Paper No. 107759:65
- Avila A, Forni G, Safaee P (2021b) Quantitative weak mixing for interval exchange transformations. <https://arxiv.org/abs/2105.10547>
- Avila A, Gouëzel S, Yoccoz J-C (2006) Exponential mixing for the Teichmüller flow. *Publ Math Inst Hautes Études Sci* 104:143–211
- Avila A, Leguil M (2018) Weak mixing properties of interval exchange transformations & translation flows. *Bull Soc Math France* 146(2):391–426
- Avila A, Viana M (2007) Simplicity of Lyapunov spectra: proof of the Zorich-Kontsevich conjecture. *Acta Math* 198(1):1–56
- Berk P, Frączek K (2015) On special flows over IETs that are not isomorphic to their inverses. *Discrete Contin Dyn Syst* 35(3):829–855
- Berk P, Frączek K, de la Rue T (2020) On typicality of translation flows which are disjoint with their inverse. *J Inst Math Jussieu* 19(5):1677–1737
- Berk P, Kanigowski A (2021) Spectral disjointness of rescalings of some surface flows. *J Lond Math Soc* (2) 103(3):901–942
- Blohin AA (1972) Smooth ergodic flows on surfaces. *Trudy Moskov Mat Obšč* 27:113–128
- Boshernitzan M, Nogueira A (2004) Generalized eigenfunctions of interval exchange maps. *Ergod Theory Dyn Syst* 24(3):697–705
- Bufetov A, Sinai YG, Ulcigrai C (2006) A condition for continuous spectrum of an interval exchange transformation. In: *Representation theory, dynamical systems, and asymptotic combinatorics*, American Mathematical Society Translations: Series 2, vol 217. American Mathematical Society, Providence, RI, pp 23–35
- Bufetov AI (2014) Limit theorems for translation flows. *Ann Math* 179(2):431–499
- Bufetov AI, Solomyak B (2018) The Hölder property for the spectrum of translation flows in genus two. *Israel J Math* 223(1):205–259
- Bufetov AI, Solomyak B (2020) A spectral cocycle for substitution systems and translation flows. *J Anal Math* 141(1):165–205
- Bufetov AI, Solomyak B (2021) Hölder regularity for the spectrum of translation flows. *J Éc Polytech Math* 8: 279–310
- Chaika J (2012) Every ergodic transformation is disjoint from almost every interval exchange transformation. *Ann Math* (2) 175(1):237–253
- Chaika J, Eskin A (2021) Self-joinings for 3-IETs. *J Eur Math Soc (JEMS)* 23(8):2707–2731
- Chaika J, Frączek K, Kanigowski A, Ulcigrai C (2021) Singularity of the spectrum for smooth area-preserving flows in genus two and translation surfaces well approximated by cylinders. *Commun Math Phys* 381(3):1369–1407
- Chaika J, Masur H (2015) There exists an interval exchange with a non-ergodic generic measure. *J Mod Dyn* 9:289–304
- Chaika J, Masur H (2020) The set of non-uniquely ergodic d-IETs has Hausdorff codimension $1/2$. *Invent Math* 222(3):749–832
- Chaika J, Robertson D (2019) Ergodicity of skew products over linearly recurrent IETs. *J Lond Math Soc* 100(1): 223–248
- Chaika J, Weiss B (2022) The horocycle flow on the moduli space of translation surfaces, ICM 2022 Proceedings. EMS Press
- Chaika J, Wright A (2019) A smooth mixing flow on a surface with nondegenerate fixed points. *J Am Math Soc* 32(1):81–117
- Cheung Y (2003) Hausdorff dimension of the set of non-ergodic directions. *Ann Math* 158(2):661–678. With an appendix by M. Boshernitzan
- Cheung Y, Hubert P, Masur H (2011) Dichotomy for the Hausdorff dimension of the set of nonergodic directions. *Invent Math* 183(2):337–383
- Conze J-P, Frączek K (2011) Cocycles over interval exchange transformations and multivalued Hamiltonian flows. *Adv Math* 226(5):4373–4428
- Danilenko AI, Ryzhikov VV (2012) On self-similarities of ergodic flows. *Proc Lond Math Soc* (3) 104(3):431–454
- de la Rue T (2020) Joinings in ergodic theory. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 1–20
- del Junco A (1981) Disjointness of measure-preserving transformations, minimal self-joinings and category. In: *Ergodic theory and dynamical systems, I* (College Park, Md., 1979–80), *Progress in Mathematics*, vol 10. Birkhäuser, Boston, MA, pp 81–89
- del Junco A (1983) A family of counterexamples in ergodic theory. *Israel J Math* 44(2):160–188

- Dong C, Kanigowski A (2020) Rigidity of a class of smooth singular flows on $\mathbb{T}^2$ *J Mod Dyn* 16:37–57
- Fayad B (2001) Polynomial decay of correlations for a class of smooth flows on the two torus. *Bull Soc Math France* 129(4):487–503
- Fayad B, Forni G, Kanigowski A (2021) Lebesgue spectrum of countable multiplicity for conservative flows on the torus. *J Am Math Soc* 34(3):747–813
- Fayad B, Kanigowski A (2016) Multiple mixing for a class of conservative surface flows. *Invent Math* 203(2):555–614
- Fayad B, Krikorian R (2018) Some questions around quasiperiodic dynamics. In: *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018, Invited lectures, vol III*. World Sci. Publ, Hackensack, NJ, pp 1909–1932
- Fayad B, Lemańczyk M (2006) On the ergodicity of cylindrical transformations given by the logarithm. *Mosc Math J* 6(4):657–672, 771–772
- Fayad BR (2002) Analytic mixing reparametrizations of irrational flows. *Ergod Theory Dyn Syst* 22(2):437–468
- Ferenczi S, Holton C, Zamboni LQ (2005) Joinings of three-interval exchange transformations. *Ergod Theory Dyn Syst* 25(2):483–502
- Ferenczi S, Zamboni LQ (2011) Eigenvalues and simplicity of interval exchange transformations. *Ann Sci Éc Norm Supér* (4) 44(3):361–392
- Forni G (1997) Solutions of the cohomological equation for area-preserving flows on compact surfaces of higher genus. *Ann Math* 146(2):295–344
- Forni G (2002) Deviation of ergodic averages for area-preserving flows on surfaces of higher genus. *Ann Math* 155(1):1–103
- Forni G (2021) Sobolev regularity of solutions of the cohomological equation. *Ergod Theory Dyn Syst* 41(3):685–789
- Forni G (2022) Twisted translation flows and effective weak mixing. *J Eur Math Soc (JEMS)* 24(2022), no. 12, 4225–4276
- Forni G, Marmi S, Matheus C (2017) Cohomological equation and local conjugacy class of diophantine interval exchange maps to appear in *Proc. Amer. Math. Soc.*, <https://doi.org/10.1090/proc/14538>
- Forni G, Matheus C (2014) Introduction to Teichmüller theory and its applications to dynamics of interval exchange transformations, flows on surfaces and billiards. *J Mod Dyn* 8(3–4):271–436
- Forni G, Ulcigrai C (2012) Time-changes of horocycle flows. *J Mod Dyn* 6(2):251–273
- Fox RH, Kershner RB (1936) Concerning the transitive properties of geodesics on a rational polyhedron. *Duke Math J* 2(1):147–150
- Frączek K, Lemańczyk M (2003) On symmetric logarithm and some old examples in smooth ergodic theory. *Fundam Math* 180(3):241–255
- Frączek K, Lemańczyk M (2004) A class of special flows over irrational rotations which is disjoint from mixing flows. *Ergod Theory Dyn Syst* 24:1083–1095
- Frączek K, Lemańczyk M (2005) On disjointness properties of some smooth flows. *Fundam Math* 185(2):117–142
- Frączek K, Lemańczyk M (2006) On mild mixing of special flows over irrational rotations under piecewise smooth functions. *Ergod Theory Dyn Syst* 26(3):719–738
- Frączek K, Lemańczyk M (2010) Ratner’s property and mild mixing for special flows over two-dimensional rotations. *J Mod Dyn* 4:609–635
- Frączek K (2009) Density of mild mixing property for vertical flows of abelian differentials. *Proc Am Math Soc* 137(12):4229–4142
- Frączek K, Hubert P (2018) Recurrence and non-ergodicity in generalized wind-tree models. *Math Nachr* 291(11–12):1686–1711
- Frączek K, Kim M (2021) New phenomena in deviation of Birkhoff integrals for locally Hamiltonian flows. *arXiv:2112.13030*
- Frączek K, Kułaga-Przymus J, Lemańczyk M (2014) Non-reversibility and self-joinings of higher orders for ergodic flows. *J Anal Math* 122:163–227
- Frączek K, Lemańczyk M (2009a) On the self-similarity problem for ergodic flows. *Proc Lond Math Soc* 99(3):658–696
- Frączek K, Lemańczyk M (2009b) Smooth singular flows in dimension 2 with the minimal self-joining property. *Monatsh Math* 156(1):11–45
- Frączek K, Lemańczyk M, Lesigne E (2007) Mild mixing property for special flows under piecewise constant functions. *Discrete Contin Dyn Syst* 19(4):691–710
- Frączek K, Ulcigrai C (2012) Ergodic properties of infinite extensions of area-preserving flows. *Math Ann* 354(4):1289–1367
- Frączek K, Ulcigrai C (2014) Non-ergodic $\mathbb{Z}$ -periodic billiards and infinite translation surfaces. *Invent Math* 197(2):241–298
- Frączek K, Ulcigrai C (2021) On the asymptotic growth of Birkhoff integrals for locally Hamiltonian flows and ergodicity of their extensions. *arXiv:2112.05939*
- Furstenberg H (1967) Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math Syst Theory* 1:1–49
- Ghazouani S (2021) Local rigidity for periodic generalised interval exchange transformations. *Invent Math* 226(2):467–520
- Ghazouani S, Ulcigrai C (2021) A priori bounds for GIETs, affine shadows and rigidity of foliations in genus 2. Preprint *arXiv:2106.03529*
- Halmos PR (1960) *Lectures on ergodic theory*. Chelsea Publishing Co., New York
- Hmili H (2010) Non topologically weakly mixing interval exchanges. *Discrete Contin Dyn Syst* 27(3):1079–1091
- Hooper WP (2015) The invariant measures of some infinite interval exchange maps. *Geom Topol* 19(4):1895–2038
- Host B (1991) Mixing of all orders and pairwise independent joinings of systems with singular spectrum. *Israel J Math* 76(3):289–298

- Hubert P, Weiss B (2013) Ergodicity for infinite periodic translation surfaces. *Compos Math* 149(8):1364–1380
- Kanigowski A (2015) Ratner's property for special flows over irrational rotations under functions of bounded variation. *Ergod Theory Dyn Syst* 35(3):915–934
- Kanigowski A, Kułaga-Przymus J (2016) Ratner's property and mild mixing for smooth flows on surfaces. *Ergod Theory Dyn Syst* 36(8):2512–2537
- Kanigowski A, Kułaga-Przymus J, Ulcigrai C (2019) Multiple mixing and parabolic divergence in smooth area-preserving flows on higher genus surfaces. *J Eur Math Soc (JEMS)* 21(12):3797–3855
- Kanigowski A, Lemańczyk M (2020) Spectral theory of dynamical systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 1–40
- Kanigowski A, Lemańczyk M, Ulcigrai C (2020) On disjointness properties of some parabolic flows. *Invent Math* 221(1):1–111
- Katok A (2001) Cocycles, cohomology and combinatorial constructions in ergodic theory. In: *Smooth ergodic theory and its applications* (Seattle, WA, 1999), *Proceedings of Symposia in Pure Mathematics*, vol 69. American Mathematical Society, Providence, RI, pp 107–173. In collaboration with E. A. Robinson, Jr
- Katok A, Thouvenot J-P (2006) Spectral properties and combinatorial constructions in ergodic theory. In: *Handbook of dynamical systems*, vol 1B. Elsevier B. V., Amsterdam, pp 649–743
- Katok AB (1973) Invariant measures of flows on oriented surfaces. *Sov Math Dokl* 14:1104–1108
- Katok AB (1980) Interval exchange transformations and some special flows are not mixing. *Israel J Math* 35(4):301–310
- Katok AB, Stepin AM (1966) Approximation of ergodic dynamical systems by periodic transformations. *Dokl Akad Nauk SSSR* 171:1268–1271
- Katok AB, Stepin AM (1967) Approximations in ergodic theory. *Uspehi Mat Nauk* 137(5):81–106
- Keane M (1975) Interval exchange transformations. *Math Z* 141:25–31
- Keane M (1977) Non-ergodic interval exchange transformations. *Israel J Math* 26(2):188–196
- Kerckhoff S, Masur H, Smillie J (1986) Ergodicity of billiard flows and quadratic differentials. *Ann Math* (2) 124(2):293–311
- Keynes HB, Newton D (1976) A “minimal”, nonuniquely ergodic interval exchange transformation. *Math Z* 148(2):101–105
- Kočergin AV (1972) The absence of mixing in special flows over a rotation of the circle and in flows on a two-dimensional torus. *Dokl Akad Nauk SSSR* 205:512–518. (Translated in: *Sov Math Dokl* 13:949–952, 1972)
- Kočergin AV (1975) Mixing in special flows over a shifting of segments and in smooth flows on surfaces. *Mat Sb* 96(138):471–502
- Kochergin AV (2004) Some generalizations of theorems on mixing flows with nondegenerate saddles on a two-dimensional torus. *Mat Sb* 195(9):19–36
- Kochergin AV (2007) Nondegenerate saddles, and absence of mixing. II. *Mat Zametki* 81(1):145–148
- Kontsevich M (1997) Lyapunov exponents and Hodge theory. In: *The mathematical beauty of physics* (Saclay, 1996), *Advanced Series in Mathematical Physics*, vol 24. World Scientific Publishing, River Edge, NJ, pp 318–332
- Kontsevich M, Zorich A (1997) Lyapunov exponents and Hodge theory. *arXiv:hep-th/9701164*
- Kočergin AV (1976) Nondegenerate saddles, and the absence of mixing. *Mat Zametki* 19(3):453–468
- Kułaga J (2012) On the self-similarity problem for smooth flows on orientable surfaces. *Ergod Theory Dyn Syst* 32(5):1615–1660
- Kułaga-Przymus J, Lemańczyk M (2020) Sarnak's conjecture from the ergodic theory point of view. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 1–19
- Lanneau E, Marmi S, Skripchenko A (2021) Cohomological equations for linear involutions. *Dyn Syst* 36(2):292–304
- Lemańczyk M (2000) Sur l'absence de mélange pour des flots spéciaux au-dessus d'une rotation irrationnelle. *Colloq Math* 84–85:29–41
- Levitt G (1983) Feuilletages des surfaces. Thèse
- Málaga Sabogal A, Troubetzkoy S (2016) Ergodicity of the Ehrenfest wind-tree model. *C R Math Acad Sci Paris* 354(10):1032–1036
- Marmi S, Moussa P, Yoccoz J-C (2005) The cohomological equation for Roth-type interval exchange maps. *J Am Math Soc* 18(4):823–872. (electronic)
- Marmi S, Moussa P, Yoccoz J-C (2012) Linearization of generalized interval exchange maps. *Ann Math* (2) 176(3):1583–1646
- Marmi S, Ulcigrai C, Yoccoz J-C (2020) On Roth type conditions, duality and central Birkhoff sums for I.E.M. *Astérisque*, (416, *Quelques aspects de la théorie des systèmes dynamiques: un hommage à Jean-Christophe Yoccoz*, II), pp. 65–132
- Marmi S, Yoccoz J-C (2016) Hölder regularity of the solutions of the cohomological equation for Roth type interval exchange maps. *Comm Math Phys* 344(1):117–139
- Masur H (1982) Interval exchange transformations and measured foliations. *Ann Math* 115:169–200
- Masur H (1992) Hausdorff dimension of the set of non-ergodic foliations of a quadratic differential. *Duke Math J* 66(3):387–442
- Masur H, Smillie J (1991) Hausdorff dimension of sets of nonergodic measured foliations. *Ann Math* (2) 134(3):455–543
- Masur H, Tabachnikov S (2002) Rational billiards and flat structures. In: *Handbook of dynamical systems*, vol 1A. North-Holland, Amsterdam, pp 1015–1089
- Mayer A (1943) Trajectories on the closed orientable surfaces. *Rec Math [Mat Sbornik] NS* 12(54):71–84
- McMullen CT (2020) Teichmüller dynamics and unique ergodicity via currents and Hodge theory. *J Reine Angew Math* 768:39–54

- Nikolaev I, Zhuzhoma E (1999) Flows on 2-dimensional manifolds. In: *Lecture notes in mathematics*, vol 1705. Springer
- Nogueira A, Rudolph D (1997) Topological weak-mixing of interval exchange maps. *Ergod Theory Dyn Syst* 17(5):1183–1209
- Novikov SP (1982) The Hamiltonian formalism and a multivalued analogue of Morse theory. (Russian) *Uspekhi Matematicheskikh Nauk* 37(5):3–49. (Translated in: *Russ Math Surv* 37(5):1–56, 1982)
- Oseledec VI (1966) The spectrum of ergodic automorphisms. *Dokl Akad Nauk SSSR* 168:1009–1011
- Poincaré H (1987) *Les méthodes nouvelles de la mécanique céleste*. In: *Les Grands Classiques Gauthier-Villars, Librairie Scientifique et Technique Albert Blanchard*, Paris
- Quas A (2017) *Ergodicity and mixing properties*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 1–20
- Ralston D, Troubetzkoy S (2017) Residual generic ergodicity of periodic group extensions over translation surfaces. *Geom Dedicata* 187:219–239
- Ravotti D (2017) Quantitative mixing for locally Hamiltonian flows with saddle loops on compact surfaces. *Ann Henri Poincaré* 18(12):3815–3861
- Robertson D (2019) Mild mixing of certain interval-exchange transformations. *Ergod Theory Dyn Syst* 39(1):248–256
- Ryzhikov VV (1994) The absence of mixing in special flows over rearrangements of segments. *Mat Zametki* 55(6):146–149. (Translated in: *Math Notes* 55(5–6):648–650, 1994)
- Ryzhikov VV, Tuveno Z-P (2006) Disjointness, divisibility, and quasi-simplicity of measure-preserving actions. *Funktsional Anal i Prilozhen* 40(3):85–89
- Sataev EA (1975) The number of invariant measures for flows on orientable surfaces. *Izv Akad Nauk SSSR Ser Mat* 39(4):860–878
- Scheglov D (2009) Absence of mixing for smooth flows on genus two surfaces. *J Mod Dyn* 3(1):13–34
- Schmidt K (2002) Dispersing cocycles and mixing flows under functions. *Fundam Math* 173(2):191–199
- Sinai YG, Khanin KM (1992) Mixing for some classes of special flows over rotations of the circle. *Funktsional'nyi Analiz i Ego Prilozheniya* 26(3):1–21. (Translated in: *Funct Anal Appl* 26(3):155–169, 1992)
- Sinai YG, Ulcigrai C (2005) Weak mixing in interval exchange transformations of periodic type. *Lett Math Phys* 74(2):111–133
- Ulcigrai C (2022) Dynamics and ‘arithmetics’ of higher genus surface flows. EMS Press, ICM 2022 Proceedings
- Ulcigrai C (2007) Mixing of asymmetric logarithmic suspension flows over interval exchange transformations. *Ergod Theory Dyn Syst* 27(3):991–1035
- Ulcigrai C (2009) Weak mixing for logarithmic flows over interval exchange transformations. *J Mod Dyn* 3(1):35–49
- Ulcigrai C (2011) Absence of mixing in area-preserving flows on surfaces. *Ann Math* (2) 173(3):1743–1778
- Ulcigrai C (2021) Slow chaos in surface flows. *Boll Unione Mat Ital* 14(1):231–255
- Veech WA (1969) Strict ergodicity in zero dimensional dynamical systems and the Kronecker-Weyl theorem mod 2. *Trans Am Math Soc* 140:1–33
- Veech WA (1982a) A criterion for a process to be prime. *Monatsh Math* 94(4):335–341
- Veech WA (1982b) Gauss measures for transformations on the space of interval exchange maps. *Ann Math* 115:201–242
- Veech WA (1984) The metric theory of interval exchange transformations I. generic spectral properties. *Am J Math* 107(6):1331–1359
- Viana M (n.d.) Dynamics of interval exchange transformations and Teichmüller flows. Available from <http://w3.impa.br/viana>. Lecture Notes
- von Neumann J (1932) Zur Operatorenmethode in der klassischen Mechanik. *Ann Math* (2) 33(3):587–642
- Yoccoz J-C (2010) Interval exchange maps and translation surfaces. In: *Homogeneous flows, moduli spaces and arithmetic*, Clay Mathematics Proceedings, vol 10. American Mathematical Society, Providence, RI, pp 1–69
- Zemljakov AN, Katok AB (1975) Topological transitivity of billiards in polygons. *Mat Zametki* 18(2):291–300
- Zorich A (1984) S. P. Novikov’s problem of the semiclassical motion of an electron in a homogeneous magnetic field that is close to rational. *Uspekhi Mat Nauk* 39(5(239)):235–236
- Zorich A (1997) Deviation for interval exchange transformations. *Ergod Theory Dyn Syst* 17(6):1477–1499
- Zorich A (1999) How do the leaves of a closed 1-form wind around a surface? In: *Pseudoperiodic topology*, American Mathematical Society Translations: Series 2, vol 197. American Mathematical Society, Providence, RI, pp 135–178
- Zorich A (2006) Flat surfaces. In: *Frontiers in number theory, physics, and geometry, I*. Springer, Berlin, pp 437–583

Pressure and Equilibrium States in Ergodic Theory

Jean-René Chazottes¹ and Gerhard Keller²

¹Centre de Physique Théorique, CNRS/IP Paris, Palaiseau, France

²Department Mathematik, Universität Erlangen-Nürnberg, Erlangen, Germany

Article Outline

Glossary

Definition of the Subject

Introduction

Warming Up: Thermodynamic Formalism for Finite Systems

Shift Spaces, Invariant Measures, and Entropy

The Variational Principle: A Global

Characterization of Equilibrium

The Gibbs Property: A Local Characterization of Equilibrium

Examples on Shift Spaces

Examples from Differentiable Dynamics

Nonequilibrium Steady States and Entropy Production

Some Ongoing Developments and Future

Directions

Bibliography

Glossary

Dynamical system In this entry: a continuous transformation T of a compact metric space X . For each $x \in X$, the transformation T generates a trajectory $(x, Tx, T^2x, \dots)$.

Entropy In this entry: the maximal rate of information gain per time that can be achieved by coarse-grained observations on a measure-preserving dynamical system. This quantity is often denoted $h(\mu)$.

Equilibrium state In general, a given dynamical system $T : X \rightarrow X$ admits a huge number of invariant measures. Given some continuous

$\phi : X \rightarrow \mathbb{R}$ (“potential”), those invariant measures that maximize a functional of the form $F(\mu) = h(\mu) + \langle \phi, \mu \rangle$ are called “equilibrium states” for ϕ .

Ergodic theory Ergodic theory is the mathematical theory of measure-preserving dynamical systems.

Gibbs state In many cases, equilibrium states have a local structure that is determined by the local properties of the potential ϕ . They are called “Gibbs states.”

Invariant measure In this entry: a probability measure μ on X which is invariant under the transformation T , that is, for which $\langle f \circ T, \mu \rangle = \langle f, \mu \rangle$ for each continuous $f : X \rightarrow \mathbb{R}$. Here $\langle f, \mu \rangle$ is a short-hand notation for $\int_X f \, d\mu$. The triple (X, T, μ) is called a measure-preserving dynamical system.

Pressure The maximum of the functional $F(\mu)$ is denoted by $P(\phi)$ and called the “topological pressure” of ϕ , or simply the “pressure” of ϕ .

Sinai-Ruelle-Bowen measure Special equilibrium or Gibbs states that describe the statistics of the attractor of certain smooth dynamical systems.

Definition of the Subject

Gibbs and equilibrium states of one-dimensional lattice models in statistical physics play a prominent role in the statistical theory of chaotic dynamics. They first appear in the ergodic theory of certain differentiable dynamical systems, called “uniformly hyperbolic systems,” mainly Anosov and Axiom A diffeomorphisms (and flows). The central idea is to “code” the orbits of these systems into (infinite) symbolic sequences of symbols by following their history on a finite partition of their phase space. This defines a nice shift dynamical system called a subshift of finite type or a topological Markov chain. Then the construction of their “natural” invariant measures and the study of their properties are carried out at the symbolic

level by constructing certain equilibrium states in the sense of statistical mechanics which turn out to be also Gibbs states. The study of uniformly hyperbolic systems brought out several ideas and techniques which turned out to be extremely fruitful for the study of more general systems. Let us mention the concept of Markov partition and its avatars, the very important notion of SRB measure (after Sinai, Ruelle, and Bowen) and transfer operators. Recently, there was a revival of interest in Axiom A systems as models to understand nonequilibrium statistical mechanics.

Introduction

Our goal is to present the basic results on one-dimensional Gibbs and equilibrium states viewed as special invariant measures on symbolic dynamical systems, and then to describe without technicalities a sample of results they allow to obtain for certain differentiable dynamical systems. We hope that this contribution will illustrate the symbiotic relationship between ergodic theory and statistical mechanics, and also information theory.

We start by putting Gibbs and equilibrium states in a general perspective. The theory of Gibbs states and equilibrium states, or Thermodynamic Formalism, is a branch of rigorous Statistical Physics. The notion of a Gibbs state dates back to R.L. Dobrushin (1968–1969) (Dobrushin 1968a, b, c, 1969) and O.E. Lanford and D. Ruelle (1969) who proposed it as a mathematical idealization of an equilibrium state of a physical system which consists of a very large number of interacting components. For a finite number of components, the foundations of statistical mechanics were already laid in the nineteenth century. There was the well-known Maxwell-Boltzmann-Gibbs formula for the equilibrium distribution of a physical system with given energy function. From the mathematical point of view, the intrinsic properties of very large objects can be made manifest by performing suitable limiting procedures. Indeed, the crucial step made in the 1960s was to define the notion of a Gibbs measure or Gibbs state for a system with an infinite number of interacting components. This was done by the familiar probabilistic idea of specifying the

interdependence structure by means of a suitable class of conditional probabilities built up according to the Maxwell-Boltzmann-Gibbs formula (Georgii 1988). Notice that Gibbs states are often called “DLR states” in honor of Dobrushin, Lanford, and Ruelle. The remarkable aspect of this construction is the fact that a Gibbs state for a given type of interaction may fail to be unique. In physical terms, this means that a system with this interaction can take several distinct equilibria. The phenomenon of non-uniqueness of a Gibbs measure can thus be interpreted as a phase transition. Therefore, the conditions under which an interaction leads to a unique or to several Gibbs measures turn out to be of central importance. While Gibbs states are defined locally by specifying certain conditional probabilities, equilibrium states are defined globally by a *variational principle*: they maximize the entropy of the system under the (linear) constraint that the mean energy is fixed. Gibbs states are always equilibrium states, but the two notions do not coincide in general. However, for a class of sufficiently regular interactions, equilibrium states are also Gibbs states.

In the effort of trying to understand phase transitions, simplified mathematical models were proposed, the most famous one being undoubtedly the Ising model. This is an example of a lattice model. The set of configurations of a lattice model is $X = A^{\mathbb{Z}^d}$, where A is a finite set, which is invariant by “spatial” translations. For the physical interpretation, X can be thought, for instance, as the set of infinite configurations of a system of spins on a crystal lattice $\mathbb{Z}^d$ and one may take $A = \{-1, +1\}$, that is, spins can take two orientations, “up” and “down.” The Ising model is defined by specifying an interaction (or potential) between spins and studying the corresponding (translation-invariant) Gibbs states. The striking phenomenon is that for $d = 1$ there is a unique Gibbs state (in fact a Markov measure) whereas if $d \geq 2$, there may be several Gibbs states, although the interaction is very simple (Georgii 1988).

Equilibrium states and Gibbs states of one-dimensional lattice models ($d = 1$) played a prominent role in understanding the ergodic properties of certain types of differentiable dynamical systems, namely uniformly hyperbolic systems, Axiom A diffeomorphisms in particular. The link

between one-dimensional lattice systems and dynamical systems is made by *symbolic dynamics*. Informally, symbolic dynamics consists of replacing the orbits of the original system by its history on a finite partition of its phase space labeled by the elements of the “alphabet” A . Therefore, each orbit of the original system is replaced by an infinite sequence of symbols, that is, by an element of the set $A^{\mathbb{Z}}$ or $A^{\mathbb{N}}$, depending on whether the map describing the dynamics is invertible or not. The action of the map on an initial condition is then easily seen to correspond to the translation (or shift) of its associated symbolic sequence. In general there is no reason to get all sequences of $A^{\mathbb{Z}}$ or $A^{\mathbb{N}}$. Instead one gets a closed invariant subset X (a subshift) which can be very complicated. For a certain class of dynamical systems the partition can be successfully chosen so as to form a *Markov partition*. In this case, the dynamical system under consideration can be coded by a *subshift of finite type* (also called a *topological Markov chain*) which is a very nice symbolic dynamical system. Then one can play the game of statistical physics: for a given continuous, real-valued function (a “potential”) on X , construct the corresponding Gibbs states and equilibrium states. If the potential is regular enough, one expects uniqueness of the Gibbs state and that it is also the unique equilibrium state for this potential. This circle of ideas – ranging from Gibbs states on finite systems over invariant measures on symbolic systems and their (Shannon-)entropy with a digression to Kolmogorov-Chaitin complexity to equilibrium states and Gibbs states on subshifts of finite type – is presented in the next four sections.

At this point it should be remembered that the objects which can actually be observed are not equilibrium states (they are measures on X) but individual symbol sequences in X , which reflect more or less the statistical properties of an equilibrium state. Indeed, most sequences reflect these properties very well, but there are also rare sequences that look quite different. Their properties are described by *large deviations principles* which are not discussed in the present article. We shall indicate some references along the way.

In sections “[Examples on Shift Spaces](#)” and “[Examples from Differentiable Dynamics](#)” we

present a selection of important examples: measure of maximal entropy, Markov measures and Hofbauer’s example of nonuniqueness of equilibrium state; uniformly expanding Markov maps of the interval, interval maps with an indifferent fixed point, Anosov diffeomorphisms and Axiom A attractors with Sinai-Ruelle-Bowen measures, and Bowen’s formula for the Hausdorff dimension of conformal repellers. As we shall see, Sinai-Ruelle-Bowen measures are the only physically observable measures and they appear naturally in the context of nonuniformly hyperbolic diffeomorphisms (Young 2002).

A revival of the interest to Anosov and Axiom A systems occurred in statistical mechanics in the 1990s. Several physical phenomena of non-equilibrium origin, like entropy production and chaotic scattering, were modeled with the help of those systems (by G. Gallavotti, P. Gaspard, D. Ruelle, and others). This new interest led to new results about old Anosov and Axiom A systems, see, for example, Chernov (2002) for a survey and references. In section “[Non-equilibrium Steady States and Entropy Production](#),” we give a very brief account of *entropy production* in the context of Anosov systems which highlights the role of *relative entropy*.

This entry is a little introduction to a vast subject in which we have tried to put forward some aspects not previously described in other expository texts. For readers willing to deepen their understanding of equilibrium and Gibbs states, there are the classic monographs by Bowen (2017) and by Ruelle (2004), the monograph by one of us (Keller 1998), and the survey article by Chernov (2002) (where Anosov and Axiom A flows are reviewed). Those texts are really complementary.

Warming Up: Thermodynamic Formalism for Finite Systems

We introduce the thermodynamic formalism in an elementary context, following Jaynes (1989). In this view, entropy, in the sense of information theory, is the central concept.

Incomplete knowledge about a system is conveniently described in terms of probability

distributions on the set of its possible states. This is particularly simple if the set of states, call it X , is finite. Then the equidistribution on X describes complete lack of knowledge, whereas a probability vector that assigns probability 1 to one single state and probability 0 to all others represents maximal information about the system. A well-established measure of the amount of uncertainty represented by a probability distribution $\nu = (\nu(x))_{x \in X}$ is its *entropy*

$$H(\nu) := - \sum_{x \in X} \nu(x) \log \nu(x),$$

which is zero if the probability is concentrated in one state and which attains its maximum value $\log|X|$ if ν is the equidistribution on X , that is, if $\nu(x) = |X|^{-1}$ for all $x \in X$. In this completely elementary context we will explore two concepts whose generalizations are central to the theory of equilibrium states in ergodic theory:

- Equilibrium distributions – defined in terms of a variational problem
- The Gibbs property of equilibrium distributions

The only mathematical prerequisite for this section are calculus and some elements from probability theory.

Equilibrium Distributions and the Gibbs Property

Suppose that a finite system can be observed through a function $U : X \rightarrow \mathbb{R}$ (an “observable”), and that we are looking for a probability distribution μ which maximizes entropy among all distributions ν with a prescribed expected value $\langle U, \nu \rangle := \sum_{x \in X} \nu(x) U(x)$ for the observable U . This means we have to solve a variational problem under constraints:

$$H(\mu) = \max \{H(\nu) : \langle U, \nu \rangle = E\}. \quad (1)$$

As the function $\nu \mapsto H(\nu)$ is strictly concave, there is a unique maximizing probability

distribution μ provided the value E can be attained at all by some $\langle U, \nu \rangle$. In order to derive an explicit formula for this μ we introduce a Lagrange multiplier $\beta \in \mathbb{R}$ and study, for each β , the unconstrained problem

$$H(\mu_\beta) + \langle \beta U, \mu_\beta \rangle = p(\beta U) := \max_{\nu} (H(\nu) + \langle \beta U, \nu \rangle). \quad (2)$$

In analogy to the convention in ergodic theory we call $p(\beta U)$ the *pressure* of βU and the maximizer μ_β the corresponding *equilibrium distribution* (synonymously *equilibrium state*).

The equilibrium distribution μ_β satisfies

$$\mu_\beta(x) = \exp(-p(\beta U) + \beta U(x)) \quad \text{for all } x \in X \quad (3)$$

as an elementary calculation using Jensen’s inequality for the strictly convex function $t \mapsto -t \log t$ shows:

$$\begin{aligned} H(\nu) + \langle \beta U, \nu \rangle &= \sum_{x \in X} \nu(x) \log \frac{e^{\beta U(x)}}{\nu(x)} \\ &\leq \log \sum_{x \in X} \nu(x) \frac{e^{\beta U(x)}}{\nu(x)} \\ &= \log \sum_{x \in X} e^{\beta U(x)}, \end{aligned}$$

with equality if and only if $e^{\beta U}$ is a constant multiple of ν . The observation that $\nu = \mu_\beta$ is a maximizer proves at the same time that $p(\beta U) = \log \sum_{x \in X} e^{\beta U(x)}$.

The equality expressed in (3) is called the *Gibbs property* of μ_β , and we say that μ_β is a Gibbs distribution if we want to stress this property.

In order to solve the constrained problem (1) it remains to show that there is a unique multiplier $\beta = \beta(E)$ such that $\langle U, \mu_\beta \rangle = E$. This follows from the fact that the map $\beta \mapsto \langle U, \mu_\beta \rangle$ maps the real line monotonically onto the interval $(\min U, \max U)$ which, in turn, is a direct consequence of the formulas for the first and second derivative of $p(\beta U)$ w.r.t β :

$$\begin{aligned} \frac{dp}{d\beta} &= \langle U, \mu_\beta \rangle, \quad \frac{d^2p}{d\beta^2} \\ &= \langle U^2, \mu_\beta \rangle - \langle U, \mu_\beta \rangle^2. \end{aligned} \quad (4)$$

As the second derivative is nothing but the variance of U under μ_β , it is strictly positive (except when U is a constant function), so that $\beta \mapsto \langle U, \mu_\beta \rangle$ is indeed strictly increasing. Observe also that $dp/d\beta$ is indeed the directional derivative of $p : \mathbb{R}^{|A|} \rightarrow \mathbb{R}$ in direction U . Hence the first identity in (4) can be rephrased as: μ_β is the gradient at βU of the function p .

A similar analysis can be performed for an $\mathbb{R}^d$ -valued observable U . In that case a vector $\beta \in \mathbb{R}^d$ of Lagrange multipliers is needed to satisfy the d linear constraints.

Systems on a Finite Lattice

We now assume that the system has a lattice structure, modeling its extension in space, for instance. The system can be in different states at different positions. More specifically, let $\mathbb{L}_n = \{0, 1, \dots, n-1\}$ be a set of n positions in space, let A be a finite set of states that can be attained by the system at each of its sites, and denote by $X := A^{\mathbb{L}_n}$ the set of all configurations of states from A at positions of $\mathbb{L}_n$. It is helpful to think of X as the set of all words of length n over the alphabet A . We focus on observables U_n which are sums of many local contributions in the sense that $U_n(a_0 \dots a_{n-1}) = \sum_{i=0}^{n-1} \phi(a_i \dots a_{i+r-1})$ for some “local observable” $\phi : A^r \rightarrow \mathbb{R}$. (The index $i+r-1$ has to be taken modulo n .) In terms of ϕ the maximizing measure can be written as

$$\begin{aligned} \mu_\beta(a_0 \dots a_{n-1}) \\ = \exp \left(-nP(\beta\phi) + \beta \sum_{i=0}^{n-1} \phi(a_i \dots a_{i+r-1}) \right), \end{aligned} \quad (5)$$

where $P(\beta\phi) := n^{-1}p(\beta U_n)$. A first immediate consequence of (5) is the invariance of μ_β under a cyclic shift of its argument, namely $\mu_\beta(a_1 \dots a_{n-1} a_0) = \mu_\beta(a_0 \dots a_{n-1})$. Therefore, we can restrict the maximizations in (1) and (2) to probability distributions ν

which are invariant under cyclic translations which yields

$$\begin{aligned} P(\beta\phi) &= \max_{\nu} (n^{-1}H(\nu) + \langle \beta\phi, \nu \rangle) \\ &= n^{-1}H(\mu_\beta) + \langle \beta\phi, \mu_\beta \rangle. \end{aligned} \quad (6)$$

If the local observable ϕ depends only on one coordinate, μ_β turns out to be a product measure:

$$\mu_\beta(a_0 \dots a_{n-1}) = \prod_{i=0}^{n-1} \exp(-P(\beta\phi) + \beta\phi(a_i)).$$

Indeed, comparison with (3) shows that μ_β is the n -fold product of the probability distribution μ_β^{loc} on A that maximizes $H(\nu) + \beta\nu(\phi)$ among all distributions ν on A . It follows that $n^{-1}H(\mu_\beta) = H(\mu_\beta^{\text{loc}})$ so that (6) implies $P(\beta\phi) = p(\beta\phi)$ for observables ϕ that depend only on one coordinate.

Shift Spaces, Invariant Measures, and Entropy

We now turn to *shift dynamical systems* over a finite alphabet A .

Symbolic Dynamics

We start by fixing some notation. Let $\mathbb{N}$ denote the set $\{0, 1, 2, \dots\}$. In the sequel we need

- A finite set A (the “alphabet”),
- The set $A^{\mathbb{N}}$ of all infinite sequences over A , that is, the set of all $\underline{x} = x_0x_1\dots$ with $x_n \in A$ for all $n \in \mathbb{N}$,
- The translation (or shift) $\sigma : A^{\mathbb{N}} \rightarrow A^{\mathbb{N}}$, $(\sigma\underline{x})_n = x_{n+1}$, for all $n \in \mathbb{N}$,
- A shift invariant subset $X = \sigma(X)$ of $A^{\mathbb{N}}$. With a slight abuse of notation we denote the restriction of σ to X by σ again.

We mention two interpretations of the dynamics of σ : it can describe the evolution of a system with state space X in discrete time steps (this is the prevalent interpretation if $\sigma : X \rightarrow X$ is obtained as a symbolic representation of another dynamical system), or it can be the spatial translation of the

configuration of a system on an infinite lattice (generalizing the point of view from subsection “[Systems on a Finite Lattice](#)” above). In the latter case one usually looks at the shift on the two-sided shift space $A^{\mathbb{Z}}$, for which the theory is nearly identical.

On $A^{\mathbb{N}}$ one can define a metric d by

$$d(\underline{x}, \underline{y}) := 2^{-N(\underline{x}, \underline{y})} \quad \text{where } N(\underline{x}, \underline{y}) := \min \{k \in \mathbb{N} : x_k \neq y_k\}. \quad (7)$$

Hence $d(\underline{x}, \underline{y}) = 1$ if and only if $x_0 \neq y_0$, and $d(\underline{x}, \underline{x}) = 0$ upon agreeing that $N(\underline{x}, \underline{x}) = \infty$ and $2^{-\infty} = 0$. Equipped with this metric, $A^{\mathbb{N}}$ becomes a compact metric space and σ is easily seen to be a continuous surjection of $A^{\mathbb{N}}$. Finally, if X is a closed subset of $A^{\mathbb{N}}$, we call the restriction $\sigma : X \rightarrow X$, which is again a continuous surjection, a shift dynamical system. We remark that d generates on $A^{\mathbb{N}}$ the product topology of the discrete topology on A , just as many variants of d do. For more details ▶ “[Symbolic Dynamics](#)”. As usual, $C(X)$ denotes the space of real-valued continuous functions on X equipped with the supremum norm $\|\cdot\|_{\infty}$.

Invariant Measures

A probability distribution ν (or simply distribution) on X is a Borel probability measure on X . It is unambiguously specified by its values $\nu[a_0 \dots a_{n-1}]$ ($n \in \mathbb{N}$, $a_i \in A$) on *cylinder sets*

$$[a_0 \dots a_{n-1}] := \{\underline{x} \in X : x_i = a_i \text{ for all } i = 0, \dots, n-1\}.$$

Any bounded and measurable $f : X \rightarrow \mathbb{R}$ (in particular any $f \in C(X)$) can be integrated by any distribution ν . To stress the linearity of the integral in both, the integrand and the integrator, we use the notation

$$\langle f, \nu \rangle := \int_X f \, d\nu.$$

In probabilistic terms, $\langle f, \nu \rangle$ is the expectation of the observable f under ν . The set $\mathcal{M}(X)$ of all

probability distributions is compact in the weak topology, the coarsest topology on $\mathcal{M}(X)$ for which $\nu \mapsto \langle f, \nu \rangle$ is continuous for all $f \in C(X)$, “Measure Preserving Systems”, subsection “Existence of Invariant Measures.” (Note that in functional analysis this is called the weak-* topology.) Henceforth we will use both terms, “measure” and “distribution,” if we talk about probability distributions.

A measure ν on X is *invariant* if expectations of observables are unchanged under the shift, that is, if

$$\langle f \circ \sigma, \nu \rangle = \langle f, \nu \rangle \quad \text{for all bounded measurable } f : X \rightarrow \mathbb{R}.$$

The set of all invariant measures is denoted by $\mathcal{M}_{\sigma}(X)$. As a closed subset of $\mathcal{M}(X)$ it is compact in the weak topology. Of special importance among all invariant measures ν are the ergodic ones which can be characterized by the property that, for all bounded measurable $f : X \rightarrow \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(\sigma^k \underline{x}) = \langle f, \nu \rangle \quad (8) \quad \text{for } \nu - \text{a.e. (almostevery) } \underline{x},$$

that is, for a set of $\underline{x}$ of ν -measure one. They are the indecomposable “building blocks” of all other measures in $\mathcal{M}_{\sigma}(X)$, “Measure Preserving Systems” or ▶ “[Ergodic Theorems](#).” The almost everywhere convergence in (8) is Birkhoff’s ergodic theorem ▶ “[Ergodic Theorems](#),” the constant limit characterizes the ergodicity of ν .

Entropy of Invariant Measures

We give a brief account of the definition and basic properties of the entropy of an invariant measure ν . For details and the generalization of this concept to general dynamical systems, we refer to ▶ “[Entropy in Ergodic Theory](#)” or Katok and Hasselblatt (1995), and to Katok (2007) for an historical account.

Let $\nu \in \mathcal{M}_{\sigma}(X)$. For each $n > 0$ the cylinder probabilities $\nu[a_0 \dots a_{n-1}]$ give rise to a probability distribution on the finite set $A^{\mathbb{L}_n}$, see section “[Warming Up: Thermodynamic Formalism for Finite Systems](#),” so

$$H_n(v) := - \sum_{a_0, \dots, a_{n-1} \in A} v[a_0 \dots a_{n-1}] \log v[a_0 \dots a_{n-1}]$$

is well defined. Invariance of v guarantees that the sequence $(H_n(v))_{n>0}$ is *subadditive*, that is, $H_{k+n}(v) \leq H_k(v) + H_n(v)$, and an elementary argument shows that the limit

$$h(v) := \lim_{n \rightarrow \infty} \frac{1}{n} H_n(v) \in [0, \log |A|] \quad (9)$$

exists and equals the infimum of the sequence. We simply call it the *entropy* of v . (Note that for general subshifts X many of the cylinder sets $[a_0 \dots a_{n-1}] \subseteq X$ are empty. But, because of the continuity of the function $t \mapsto t \log t$ at $t = 0$, we may set $0 \log 0 = 0$, and, hence, this does not affect the definition of $H_n(v)$.)

The entropy $h(v)$ of an ergodic measure v can be observed along a “typical” trajectory. That is the content of the following theorem, sometimes called the “ergodic theorem of information theory” (► “Entropy in Ergodic Theory”).

Theorem (Shannon-McMillan-Breiman Theorem)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log v[x_0 \dots x_{n-1}] = -h(v) \quad \text{for } v\text{-a. e. } \underline{x}. \quad (10)$$

Observe that (9) is just the integrated version of this statement. A slightly weaker reformulation of this theorem (again for ergodic v) is known as the “asymptotic equipartition property.”

Asymptotic Equipartition Property Given (arbitrarily small) $\epsilon > 0$ and $\alpha > 0$, one can, for each sufficiently large n , partition the set A^n into a set $\mathcal{T}_n$ of typical words and a set $\mathcal{E}_n$ of exceptional words such that each $a_0 \dots a_{n-1} \in \mathcal{T}_n$ satisfies

$$e^{-n(h(v)+\alpha)} \leq v[a_0 \dots a_{n-1}] \leq e^{-n(h(v)-\alpha)} \quad (11)$$

and the total probability $\sum_{a_0 \dots a_{n-1} \in \mathcal{E}_n} v[a_0 \dots a_{n-1}]$ of the exceptional words is at most ϵ .

A Short Digression on Complexity

Kolmogorov (1983) and Chaitin (1987) introduced the concept of complexity of an infinite sequence of symbols. Very roughly it is defined as follows: First, the complexity $K(x_0 \dots x_{n-1})$ of a finite word in A^n is defined as the bit length of the shortest program that causes a suitable general purpose computer (say a PC or, for the mathematically minded reader, a Turing machine) to print out this word. Then the complexity of an infinite sequence is defined as $K(\underline{x}) := \limsup_{n \rightarrow \infty} \frac{1}{n} K(x_0 \dots x_{n-1})$. Of course, the definition of $K(x_0 \dots x_{n-1})$ depends on the particular computer, but as any two general purpose computers can be programmed to simulate each other (by some finite piece of software), the limit $K(\underline{x})$ is machine independent. It is the optimal compression factor for long initial pieces of a sequence $\underline{x}$ that still allows complete reconstruction of $\underline{x}$ by an algorithm. Brudno (1983) showed:

If $X \subseteq A^{\mathbb{N}}$ and $v \in \mathcal{M}_\sigma(X)$ is ergodic, then $K(\underline{x}) = \frac{1}{\log 2} h(v)$ for v -a. e. $\underline{x} \in X$.

Entropy as a Function of the Measure

An important technical remark for the further development of the theory is that the entropy function $h : \mathcal{M}_\sigma(X) \rightarrow [0, \infty)$ is *upper semicontinuous*. This means that all sets $\{v : h(v) \geq t\}$ with $t \in \mathbb{R}$ are closed and hence compact. In particular, upper semicontinuous functions attain their supremum. Indeed, suppose a sequence $v_k \in \mathcal{M}_\sigma(X)$ converges weakly to some $v \in \mathcal{M}_\sigma(X)$ and $h(v_k) \geq t$ for all k so that also $\frac{1}{n} H_n(v_k) \geq t$ for all n and k . As $H_n(v)$ is an expression that depends continuously on the probabilities of the finitely many cylinders $[a_0 \dots a_{n-1}]$ and as the indicator functions of these sets are continuous, $\frac{1}{n} H_n(v) = \lim_{k \rightarrow \infty} \frac{1}{n} H_n(v_k) \geq t$, hence $h(v) \geq t$ in the limit $n \rightarrow \infty$.

A word of caution seems in order: the entropy function is rarely continuous. For example, on the full shift $X = A^{\mathbb{N}}$ each invariant measure, whatever its entropy is, can be approximated in the weak topology by equidistributions on periodic orbits. But all these equidistributions have entropy zero.

The Variational Principle: A Global Characterization of Equilibrium

Usually, a dynamical systems model of a “physical” system consists of a state space and a map (or a differential equation) describing the dynamics. An invariant measure for the system is rarely given a priori. Indeed, many (if not most) dynamical systems arising in this way have uncountably many ergodic invariant measures. This limits considerably the “practical value” of Birkhoff’s ergodic theorem (8) or the Shannon-McMillan-Breiman theorem (10): not only do the limits in these theorems depend on the invariant measure ν , but also the sets of points for which the theorems guarantee almost everywhere convergence are practically disjoint for different ν and ν' in $\mathcal{M}_\sigma(X)$. Therefore, a choice of ν has to be made which reflects the original modeling intentions. We will argue in this and the next sections that a variational principle with a judiciously chosen “observable” may be a useful guideline – generalizing the observations for finite systems collected in the corresponding section above. As announced earlier we restrict again to shift dynamical systems, because they are rather universal models for many other systems.

Equilibrium States

We define the *pressure* of an observable $\phi \in C(X)$ as

$$P(\phi) := \sup\{h(\nu) + \langle \phi, \nu \rangle : \nu \in \mathcal{M}_\sigma(X)\}. \quad (12)$$

Since $\mathcal{M}_\sigma(X)$ is compact and the functional $\nu \mapsto h(\nu) + \langle \phi, \nu \rangle$ is upper semicontinuous, the supremum is attained – not necessarily at a unique measure as we will see (which is remarkably different from what happens in finite systems). Each measure ν for which the supremum is attained is called an *equilibrium state* for ϕ . Here the word “state” is used synonymously with “distribution” or “measure” – a reflection of the fact that in “well-behaved cases,” as we will see in the next section, this measure is uniquely determined by the constraint(s) under which it maximizes

entropy, and that means by the *macroscopic state* of the system. (In contrast, the word “state” was used in the above section on finite systems to designate microscopic states.)

As, for each $\nu \in \mathcal{M}_\sigma(X)$, the functional $\phi \mapsto h(\nu) + \langle \phi, \nu \rangle$ is affine on $C(X)$, the pressure functional $P : C(X) \rightarrow \mathbb{R}$, which, by definition, is the pointwise supremum of these functionals, is convex. It is therefore instructive to fit equilibrium states into the abstract framework of convex analysis (Israel 1979; Keller 1998; Moulin-Ollagnier 1985; Walters 1992). To this end recall the identities in (4) that identify, for finite systems, equilibrium states as gradients of the pressure function $p : \mathbb{R}^{|A|} \rightarrow \mathbb{R}$ and guarantee that p is twice differentiable and strictly convex. In the present setting where P is defined on the Banach space $C(X)$, differentiability and strict convexity are no more guaranteed, but one can show:

Equilibrium States as (Sub)-gradients

$\phi \in \mathcal{M}_\sigma(X)$ is an equilibrium state for ϕ if and only if μ is a subgradient (or tangent functional) for P at ϕ , i.e., if $P(\phi + \psi) - P(\phi) \geq \langle \psi, \mu \rangle$ for all $\psi \in C(X)$. In particular, ϕ has a unique equilibrium state μ if and only if P is differentiable at ϕ with gradient μ , i.e., if,

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{1}{t} (P(\phi + t\psi) - P(\phi)) \\ = \langle \psi, \mu \rangle \text{ for all } \psi \in C(X). \end{aligned} \quad (13)$$

Let us see how equilibrium states on $X = A^\mathbb{N}$ can directly be obtained from the corresponding equilibrium distributions on finite sets A^n introduced in subsection “Systems on a Finite Lattice.” Define $\phi^{(n)} : A^n \rightarrow \mathbb{R}$ by $\phi^{(n)}(a_0 \dots a_{n-1}) := \phi(a_0 \dots a_{n-1} a_0 \dots a_{n-1} \dots)$, denote by U_n the corresponding global observable on A^n , and let μ_n be the equilibrium distribution on A^n that maximizes $H(\mu) + \langle U_n, \mu \rangle$. Then all weak limit points of the “approximative equilibrium distributions” μ_n on A^n are equilibrium states on $A^\mathbb{N}$.

This can be seen as follows: Let the measure μ on $A^\mathbb{N}$ be any weak limit point of the μ_n . Then, given $\epsilon > 0$ there exists $k \in \mathbb{N}$ such that

$$\begin{aligned} h(\mu) + \langle \phi, \mu \rangle &\geq \frac{1}{k} H_k(\mu) + \langle \phi, \mu \rangle - \epsilon \\ &\geq \frac{1}{k} H_k(\mu_n) + \langle \phi^{(n)}, \mu_n \rangle - 2\epsilon \end{aligned}$$

for arbitrarily large n , because $\|\phi - \phi^{(n)}\|_\infty \rightarrow 0$ as $n \rightarrow \infty$ by construction of the $\phi^{(n)}$. As the μ_n are invariant under cyclic coordinate shifts (see subsection “[Systems on a Finite Lattice](#)”), it follows from the subadditivity of the entropy that

$$\begin{aligned} h(\mu) + \langle \phi, \mu \rangle &\geq \frac{1}{n} (H_n(\mu_n) + \langle U_n, \mu_n \rangle) - 2\epsilon \\ &\quad - \frac{k}{n} \log |A|. \end{aligned}$$

Hence, for each $v \in \mathcal{M}_\sigma(X)$,

$$\begin{aligned} h(\mu) + \langle \phi, \mu \rangle &\geq \frac{1}{n} (H_n(v) + \langle U_n, v \rangle) - 2\epsilon \\ &\quad - \frac{k}{n} \log |A| \rightarrow h(v) + \langle \phi, v \rangle - 2\epsilon \end{aligned}$$

as $n \rightarrow \infty$, and we see that μ is indeed an equilibrium state on $A^{\mathbb{N}}$.

The Variational Principle

In subsection “[Equilibrium Distributions and the Gibbs Property](#),” the pressure of a finite system was defined as a certain supremum and then identified as the logarithm of the normalizing constant for the Gibbsian representation of the corresponding equilibrium distribution. We are now going to approximate equilibrium states by suitable Gibbs distributions on finite subsets of X . As a by-product the pressure $P(\phi)$ is characterized in terms of the logarithms of the normalizing constants of these approximating distributions. Let $S_n\phi(\underline{x}) := \phi(\underline{x}) + \phi(\sigma\underline{x}) + \dots + \phi(\sigma^{n-1}\underline{x})$. From each cylinder set $[a_0 \dots a_{n-1}]$ we can pick a point $\underline{z}$ such that $S_n\phi(\underline{z})$ is the maximal value of $S_n\phi$ on this set. We denote the collection of the $|A|^n$ points we obtain in this way by E_n . Observe that E_n is not unambiguously defined, but any choice we make will do.

Theorem (Variational Principle for the Pressure)

$$\begin{aligned} P(\phi) &= \limsup_{n \rightarrow \infty} \frac{1}{n} P_n(\phi) \\ \text{where } P_n(\phi) &:= \log \sum_{\underline{z} \in E_n} e^{S_n\phi(\underline{z})}. \end{aligned} \quad (14)$$

To prove the “ $\leq$ ” direction of this identity we just have to show that $\frac{1}{n} H_n(v) + \langle \phi, v \rangle \leq \frac{1}{n} P_n(\phi)$ for each $v \in \mathcal{M}_\sigma(X)$ or, after multiplying by n , $H_n(v) + \langle S_n\phi, v \rangle \leq P_n(\phi)$. But Jensen’s inequality implies:

$$\begin{aligned} H_n(v) + \langle S_n\phi, v \rangle &\leq \sum_{a_0, \dots, a_{n-1} \in A} v[a_0 \dots a_{n-1}] \\ &\quad \log \left(\frac{\sup \left\{ e^{S_n\phi(\underline{x})} : \underline{x} \in [a_0 \dots a_{n-1}] \right\}}{v[a_0 \dots a_{n-1}]} \right) \\ &\leq \log \sum_{a_0, \dots, a_{n-1} \in A} \sup \left\{ e^{S_n\phi(\underline{x})} : \underline{x} \in [a_0 \dots a_{n-1}] \right\} \\ &= \log \sum_{\underline{z} \in E_n} e^{S_n\phi(\underline{z})} = P_n(\phi). \end{aligned}$$

For the reverse inequality consider the discrete Gibbs distributions

$$\pi_n := \sum_{\underline{z} \in E_n} \delta_{\underline{z}} \exp(-P_n(\phi) + S_n\phi(\underline{z}))$$

on the finite sets E_n , where $\delta_{\underline{z}}$ denotes the unit point mass in $\underline{z}$. One might be tempted to think that all weak limit points of the measures π_n are already equilibrium states. But this need not be the case because there is no good reason that these limits are shift invariant. Therefore, one forces invariance of the limits by passing to measures μ_n defined by $\langle f, \mu_n \rangle := \left\langle \frac{1}{n} \sum_{k=0}^{n-1} f \circ \sigma^k, \pi_n \right\rangle$. Weak limits of these measures are obviously shift invariant, and a more involved estimate we do not present here shows that each such weak limit μ satisfies $h(\mu) + \langle \phi, \mu \rangle \geq P(\phi)$.

We note that the same arguments work for any other sequence of sets E_n which contain exactly one point from each cylinder. So there are many ways to approximate equilibrium states, and if there are more than one equilibrium state, there is generally no guarantee that the limit is always the same.

Nonuniqueness of Equilibrium States: An Example

Before we turn to sufficient conditions for the uniqueness of equilibrium states in the next section, we present one of the simplest nontrivial examples for nonuniqueness of equilibrium states. Motivated by the so-called Fisher-Felderhof drop-let model of condensation in statistical mechanics (Fisher and Felderhof 1970; Fisher 1967), Hofbauer (1977) studies an observable ϕ on $X = \{0, 1\}^{\mathbb{N}}$ defined as follows: Let (a_k) be a sequence of negative real numbers with $\lim_{k \rightarrow \infty} a_k = 0$. Set $s_k := a_0 + \dots + a_k$. For $k \geq 1$ denote $M_k := \{\underline{x} \in X : x_0 = \dots = x_{k-1} = 1, x_k = 0\}$ and $M_0 := \{\underline{x} \in X : x_0 = 0\}$, and define

$$\phi(\underline{x}) := a_k \quad \text{for } \underline{x} \in M_k \quad \text{and} \quad \phi(11\dots) = 0.$$

Then $\phi : X \rightarrow \mathbb{R}$ is continuous, so that there exists at least one equilibrium state for ϕ . Hofbauer proves that there is more than one equilibrium state if and only if $\sum_{k=0}^{\infty} e^{s_k} = 1$ and $\sum_{k=0}^{\infty} (k+1)e^{s_k} < \infty$. In that case $P(\phi) = 0$, so one of these equilibrium states is the unit mass $\delta_{11\dots}$, and we denote the other equilibrium state by μ_1 , so $h(\mu_1) + \langle \phi, \mu_1 \rangle = 0$. In view of (13) the pressure function is not differentiable at ϕ .

What does the pressure function $\beta \mapsto P(\beta\phi)$ look like? As $h(\delta_{11\dots}) + \langle \beta\phi, \delta_{11\dots} \rangle = 0$ for all β , $P(\beta\phi) \geq 0$ for all β . Observe now that $\phi(\underline{x}) \leq 0$ with equality only for $\underline{x} = 11\dots$. This implies that $\langle \phi, \mu \rangle < 0$ for all $\mu \in \mathcal{M}_\sigma(X)$ different from $\delta_{11\dots}$. From this we can conclude:

- $P(\beta\phi) \leq P(\phi) = 0$ for $\beta > 1$, so $P(\beta\phi) = 0$ for $\beta \geq 1$.
- $P(\beta\phi) \geq h(\mu_1) + \langle \beta\phi, \mu_1 \rangle = h(\mu_1) + \langle \phi, \mu_1 \rangle - (1 - \beta)\langle \phi, \mu_1 \rangle = - (1 - \beta)\langle \phi, \mu_1 \rangle$.

It follows that, at $\beta = 1$, the derivative from the right of $P(\beta\phi)$ is zero, whereas the derivative from the left is at most $-\langle \phi, \mu_1 \rangle < 0$.

More on Equilibrium States

In more general dynamical systems the entropy function is not necessarily upper semicontinuous

and hence equilibrium states need not exist, that is, the supremum in (12) need not be attained by any invariant measure. A well-known sufficient property that guarantees the upper semicontinuity of the entropy function is the *expansiveness* of the system, see, for example, Ruelle (1973): a continuous transformation T of a compact metric space is *positively expansive*, if there is a constant $\gamma > 0$ such that for any two points x and y from the space there is some $n \in \mathbb{N}$ such that $T^n x$ and $T^n y$ are at least a distance γ apart. If T is a homeomorphism one says it is *expansive*, if the same holds for some $n \in \mathbb{Z}$. The previous results carry over without changes (although at the expense of more complicated proofs) to general expansive systems. The variational principle (14) holds in the very general context where T is a continuous action of $\mathbb{Z}_+^d$ on a compact Hausdorff space X . This was proved in Misiurewicz (1976) in a simple and elegant way. In the monograph (Moulin-Ollagnier 1985) it is extended to amenable group actions.

The Gibbs Property: A Local Characterization of Equilibrium

In this section we are going to see that, for a sufficiently regular potential ϕ on a topologically mixing subshift of finite type, one has a unique equilibrium state which has the “Gibbs property.” This property generalizes formula (5) that we derived for finite lattices. Subshifts of finite type are the symbolic models for Axiom A diffeomorphisms, as we shall see later on.

Subshifts of Finite Type

We start by recalling what is a subshift of finite type and refer the reader to ► “Symbolic Dynamics” or Lind and Marcus (1995) for more details. Given a “transition matrix” $M = (M_{ab})_{a,b \in A}$ whose entries are 0’s or 1’s, one can define a subshift X_M as the set of all sequences $\underline{x} \in A^{\mathbb{N}}$ such that $M_{x_i x_{i+1}} = 1$ for all $i \in \mathbb{N}$. This is called a subshift of finite type or a topological Markov chain. We assume that there exists some integer p_0 such that M^{p_0} has strictly positive entries for all

$p \geq p_0$. This means that M is irreducible and aperiodic. This property is equivalent to the property that the subshift of finite type is topologically mixing. A general subshift of finite type admits a decomposition into a finite union of transitive sets, each of which being a union of cyclically permuted sets on which the appropriate iterate is topologically mixing. In other words, topologically mixing subshifts of finite type are the building blocks of subshifts of finite type.

The Gibbs Property for a Class of Regular Potentials

The class of regular potentials we consider is that of “summable variations.” We denote by $\text{var}_k(\phi)$ the modulus of continuity of ϕ on cylinders of length $k \geq 1$, that is,

$$\text{var}_k(\phi) := \sup \left\{ \left| \phi(\underline{x}) - \phi(\underline{y}) \right| : \underline{x} \in [y_0 \dots y_{k-1}] \right\}.$$

If $\text{var}_k(\phi) \rightarrow 0$ as $k \rightarrow \infty$, this means that ϕ is (uniformly) continuous with respect to the distance (7). We impose the stronger condition

$$\sum_{k=1}^{\infty} \text{var}_k(\phi) < \infty. \quad (15)$$

We can now state the main result of this section.

The Gibbs state of a summable potential Let X_M be a topologically mixing subshift of finite type. Given a potential $\phi : X_M \rightarrow \mathbb{R}$ satisfying the summability condition (15), there is a (probability) measure μ_ϕ supported on X_M , that we call a Gibbs state. It is the unique σ -invariant measure which satisfies the following property:

There exists a constant $C > 0$ such that, for all $\underline{x} \in X_M$ and for all $n \geq 1$,

$$\begin{aligned} C^{-1} &\leq \frac{\mu_\phi[x_0 \dots x_{n-1}]}{\exp(S_n \phi(\underline{x}) - nP(\phi))} \\ &\leq C. \text{ (“Gibbs property”)} \end{aligned} \quad (16)$$

Moreover, the Gibbs state μ_ϕ is ergodic and is also the unique equilibrium state of ϕ , that is, the unique invariant measure for which the supremum in (12) is attained.

We now make several comments on this theorem.

- The Gibbs property (16) gives a *uniform control* of the measure of *all cylinders* in terms of their “energy.” This strengthens considerably the asymptotic equipartition property (11) that we recover if we restrict (16) to the set of μ_ϕ measure 1 where Birkhoff’s ergodic Theorem (8) applies, and use the identity $\langle \phi, \mu_\phi \rangle - P(\phi) = -h(\mu_\phi)$.
- Gibbs measures on topologically mixing subshifts of finite type are ergodic (and actually mixing in a strong sense) as can be inferred from Ruelle’s Perron-Frobenius Theorem see the next subsection.
- Suppose that there is another invariant measure μ' satisfying (16), possibly with a constant C' different from C . It is easy to verify that $\mu' = f\mu$ for some μ -integrable function f by using (16) and the Radon-Nikodym Theorem. Shift invariance imposes that, μ -a. e., $f = f \circ \sigma$. Then the ergodicity of μ implies that f is a constant μ -a. e., thus $\mu' = \mu$; see Bowen (2017).
- One could define a Gibbs state by saying that it is an invariant measure μ satisfying (16) for a given continuous potential ϕ . If one does so, it is simple to verify that such a μ must also be an equilibrium state. Indeed, using (16), one can deduce that $\langle \phi, \mu \rangle + h(\mu) \geq P(\phi)$. The converse need not be true in general, see subsection “[More on Hofbauer’s Example](#)” below. But the summability condition (15) is indeed sufficient for the coincidence of Gibbs and equilibrium states. A proof of this fact can be found in Ruelle (2004) or Keller (1998).

Ruelle’s Perron-Frobenius Theorem

The powerful tool behind the theorem in the previous subsection is a far-reaching generalization of the classical Perron-Frobenius theorem for irreducible matrices. Instead of a matrix, one introduces the so-called transfer operator, also called the “Perron-Frobenius operator” or “Ruelle’s operator,” which acts on a suitable Banach space of observables. It is D. Ruelle (1968) who first introduced this operator in the context of one-dimensional lattice gases with exponentially decaying interactions. In our context, this

corresponds to Hölder continuous potentials: these are potentials satisfying $\text{var}_k(\phi) \leq c\theta^k$ for some $c > 0$ and $\theta \in (0, 1)$. A proof of “Ruelle’s Perron-Frobenius Theorem” can be found in Bowen (1974b, 2017). It was then extended to potentials with summable variations in Walters (1975). We refer to the book of V. Baladi (2000) for a comprehensive account on transfer operators in dynamical systems.

We content ourselves to define the transfer operator and state Ruelle’s Perron-Frobenius Theorem. Let $\mathcal{L} : C(X_M) \rightarrow C(X_M)$ be defined by

$$\begin{aligned} (\mathcal{L}f)(\underline{x}) &:= \sum_{\underline{y} \in \sigma^{-1}\underline{x}} e^{\phi(\underline{y})} f(\underline{y}) \\ &= \sum_{a \in A: M(a, x_0)=1} e^{\phi(a\underline{x})} f(a\underline{x}). \end{aligned}$$

(Obviously, $a\underline{x} := ax_0x_1 \dots$)

Theorem (Ruelle’s Perron-Frobenius Theorem) Let X_M be a topologically mixing subshift of finite type. Let ϕ satisfy condition (15). There exist a number $\lambda > 0$, $h \in C(X_M)$, and $\nu \in \mathcal{M}(X)$ such that $h > 0$, $\langle h, \nu \rangle = 1$, $\mathcal{L}h = \lambda h$, $\mathcal{L}^* \nu = \lambda \nu$, where $\mathcal{L}^*$ is the dual of $\mathcal{L}$. Moreover, for all $f \in C(X_M)$,

$$\|\lambda^{-n} \mathcal{L}^n f - \langle f, \nu \rangle \cdot h\|_\infty \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

By using this theorem, one can show that $\mu_\phi := h\nu$ satisfies (16) and $\lambda = e^{P(\phi)}$.

Let us remark that for potentials which are such that $\phi(\underline{x}) = \phi(x_0, x_1)$ (i.e., potentials constant on cylinders of length 2), $\mathcal{L}$ can be identified with a $|A| \times |A|$ matrix and the previous theorem boils down to the classical Perron-Frobenius theorem for irreducible aperiodic matrices (Seneta 2006). The corresponding Gibbs states are nothing but Markov chains with state space A (Chap. 3 in Georgii 1988). We shall take another point of view below (subsection “Markov Chains Over Finite Alphabets”).

Relative Entropy

We now define the relative entropy of an invariant measure $\nu \in \mathcal{M}_\sigma(X_M)$ given a Gibbs state μ_ϕ as follows. We first define

$$\begin{aligned} H_n(\nu|\mu_\phi) &:= \sum_{a_0, \dots, a_{n-1} \in A} \nu[a_0 \dots a_{n-1}] \log \frac{\nu[a_0 \dots a_{n-1}]}{\mu_\phi[a_0 \dots a_{n-1}]} \end{aligned} \quad (17)$$

with the convention $0 \log(0/0) = 0$. Now the relative entropy of ν given μ_ϕ is defined as

$$h(\nu|\mu_\phi) := \limsup_{n \rightarrow \infty} \frac{1}{n} H_n(\nu|\mu_\phi).$$

(By applying Jensen’s inequality, one verifies that $h(\nu|\mu_\phi) \geq 0$.) In fact the limit exists and can be computed quite easily using (16):

$$h(\nu|\mu_\phi) = P(\phi) - \langle \phi, \nu \rangle - h(\nu). \quad (18)$$

To prove this formula, we first make the following observation. It can be easily verified that the inequalities in (16) remain the same when $S_n \phi$ is replaced by the “locally averaged” energy $\tilde{\phi}_n := (\nu[x_0 \dots x_{n-1}])^{-1} \int_{[x_0 \dots x_{n-1}]} S_n \phi(\underline{y}) d\nu(\underline{y})$ for any cylinder with $\nu[x_0 \dots x_{n-1}] > 0$. Cylinders with ν measure zero do not contribute to the sum in (17).

We can now write that

$$\begin{aligned} & -\frac{1}{n} \log C \\ & \leq -\frac{1}{n} H_n(\nu|\mu_\phi) + \left(P(\phi) - \frac{1}{n} \langle S_n \phi, \nu \rangle - \frac{1}{n} H_n(\nu) \right) \\ & \leq \frac{1}{n} \log C. \end{aligned}$$

To finish we use that $\langle S_n \phi, \nu \rangle = n \langle \phi, \nu \rangle$ (by the invariance of ν) and we apply (9) to obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} H_n(\nu|\mu_\phi) &= P(\phi) - \langle \phi, \nu \rangle - \lim_{n \rightarrow \infty} \frac{1}{n} H_n(\nu) \\ &= P(\phi) - \langle \phi, \nu \rangle - h(\nu) \end{aligned}$$

which proves (18).

The variational principle revisited We can reformulate the variational principle in the case of a potential satisfying the summability condition (15):

$$h(\nu|\mu_\phi) = 0 \quad \text{if and only if } \nu = \mu_\phi, \quad (19)$$

that is, given μ_ϕ , the relative entropy $h(\cdot|\mu_\phi)$, as a function on $\mathcal{M}_\sigma(X_M)$, attains its minimum only at μ_ϕ .

Indeed, by (18) we have $h(v|\mu_\phi) = P(\phi) - \langle \phi, v \rangle - h(v)$. We now use (12) and the fact that μ_ϕ is the unique equilibrium state of ϕ to conclude.

More Properties of Gibbs States

Gibbs states enjoy very good statistical properties. Let us mention only a few. They satisfy the “Bernoulli property,” a very strong qualitative mixing condition (Bowen 1974b, 2017; Walters 1975). The sequence of random variables $(f \circ \sigma^n)_n$ satisfies the central limit theorem (Chernov 2002; Coelho and Parry 1990; Pollicott 2000) and a large deviation principle if f is Hölder continuous (Eizenberg et al. 1994; Keller 1998; Kifer 1990; Young 1990). Let us emphasize the central role played by relative entropy in large deviations. (The deep link between thermodynamics and large deviations is described in Lewis and Pfister (1995) in a much more general context.) Finally, the so-called multifractal analysis can be performed for Gibbs states, see, for example, Pesin and Weiss (1997).

Examples on Shift Spaces

Measure of Maximal Entropy and Periodic Points

If the observable ϕ is constant zero, an equilibrium state simply maximizes the entropy. It is called *measure of maximal entropy*. The quantity $P(0) = \sup\{h(v) : v \in \mathcal{M}_\sigma(X)\}$ is called the *topological entropy* of the subshift $\sigma : X \rightarrow X$. When X is a subshift of finite type X_M with irreducible and aperiodic transition matrix M , there is a unique measure of maximal entropy, see, for example, Lind and Marcus (1995). As a Gibbs state it satisfies (16). By summing over all cylinders $[x_0 \dots x_{n-1}]$ allowed by M , it is easy to see that the topological entropy $P(0)$ is the asymptotic exponential growth rate of the number of sequences of length n that can occur as initial segments of points in X_M . This is obviously identical to the logarithm of the largest eigenvalue of the transition matrix M .

It is not difficult to verify that the total number of periodic sequences of period n equals the trace of the matrix M^n , that is, we have the formula

$$\text{Card}\{\underline{x} \in X_M : \sigma^n \underline{x} = \underline{x}\} = \text{tr}(M^n) = \sum_{i=1}^m \lambda_i^n,$$

where $\lambda_1, \dots, \lambda_m$ are all the eigenvalues of M . Asymptotically, of course, $\text{card}\{\underline{x} \in X_M : \sigma^n \underline{x} = \underline{x}\} = e^{nP(0)} + O(|\lambda'|^n)$, where λ' is the second largest (in absolute value) eigenvalue of M .

The measure of maximal entropy, call it μ_0 , describes the distribution of periodic points in X_M : one can prove (Bowen 1974a; Katok and Hasselblatt 1995) that for any cylinder $B \subset X_M$

$$\lim_{n \rightarrow \infty} \frac{\text{Card}\{\underline{x} \in B : \sigma^n \underline{x} = \underline{x}\}}{\text{Card}\{\underline{x} \in X_M : \sigma^n \underline{x} = \underline{x}\}} = \mu_0(B).$$

In other words, the finite atomic measures that assign equal weights $1/\text{Card}\{\underline{x} \in X_M : \sigma^n \underline{x} = \underline{x}\}$ to each periodic point in X_M with period n weakly converges to μ_0 , as $n \rightarrow \infty$. Each such measure has zero entropy while $h(\mu_0) = P(0) > 0$, so the entropy is not continuous on the space of invariant measures. It is, however, upper-semicontinuous (see subsection “Entropy as a Function of the Measure”).

In fact, it is possible to approximate any Gibbs state μ_ϕ on X_M in a similar way by finite atomic measures on periodic orbits, if one assigns weights properly (see, e.g., Theorem 20.3.7 in Katok and Hasselblatt (1995)).

Markov Chains Over Finite Alphabets

Let $Q = (q_{a,b})_{a,b \in A}$ be an irreducible stochastic matrix over the finite alphabet A . It is well known (see, e.g., Seneta 2006) that there exists a unique probability vector π on A that defines a stationary Markov measure ν_Q on $X = A^{\mathbb{N}}$ by $\nu_Q[a_0 \dots a_{n-1}] = \pi_{a_0} q_{a_0 a_1} \dots q_{a_{n-2} a_{n-1}}$. We are going to identify ν_Q as the *unique Gibbs distribution* $\mu \in \mathcal{M}_\sigma(X)$ that maximizes entropy under the constraints $\mu[ab] = \mu[a]q_{ab}$, that is, $\langle \phi^{ab}, \mu \rangle = 0$ ($a, b \in A$), where $\phi^{ab} := \mathbb{1}_{[ab]} - q_{ab} \mathbb{1}_{[a]}$.

Indeed, as μ is a Gibbs measure, there are $\beta_{ab} \in \mathbb{R}$ ($a, b \in A$) and constants $P \in \mathbb{R}$, $C > 0$ such that

$$\begin{aligned} C^{-1} &\leq \frac{\mu[x_0 \dots x_{n-1}]}{\exp\left(\sum_{a,b \in A} \beta_{ab} \phi_n^{ab}(\underline{x}) - nP\right)} \\ &\leq C \end{aligned} \quad (20)$$

for all $\underline{x} \in A^{\mathbb{N}}$ and all $n \in \mathbb{N}$. Let $r_{ab} := \exp(\beta_{ab} - \sum_{b' \in A} \beta_{ab'} q_{ab'} - P)$. Then the denominator in (20) equals $r_{x_0 x_1} \dots r_{x_{n-2} x_{n-1}}$, and it follows that μ is equivalent to the stationary Markov measure defined by the (nonstochastic) matrix $(r_{ab})_{a,b \in A}$. As μ is ergodic, μ is this Markov measure, and as μ satisfies the linear constraints $\mu[ab] = \mu[a]q_{ab}$, we conclude that $\mu = \nu_Q$.

The Ising Chain

Here the task is to characterize all “spin chains” in $\underline{x} \in \{-1, +1\}^{\mathbb{N}}$ (or, more commonly, $\{-1, +1\}^{\mathbb{Z}}$) which are as random as possible with the constraint that two adjacent spins have a prescribed probability $p \neq \frac{1}{2}$ to be identical. With $\phi(\underline{x}) := x_0 x_1$ this is equivalent to requiring that $\underline{x}$ is typical for a Gibbs distribution $\mu_{\beta\phi}$ where $\beta = \beta(p)$ is such that $\langle \phi, \mu_{\beta\phi} \rangle = 2p - 1$. It follows that there is a constant $C > 0$ such that for each $n \in \mathbb{N}$ and any two “spin patterns” $\underline{a} = a_0 \dots a_{n-1}$ and $\underline{b} = b_0 \dots b_{n-1}$

$$\left| \log \frac{\mu_{\beta\phi}[a_0 \dots a_{n-1}]}{\mu_{\beta\phi}[b_0 \dots b_{n-1}]} - \beta(N_{\underline{a}} - N_{\underline{b}}) \right| \leq C,$$

where $N_{\underline{a}}$ and $N_{\underline{b}}$ are the numbers of identical adjacent spins in $\underline{a}$ and $\underline{b}$, respectively.

More on Hofbauer’s Example

We come back to the example described in subsection “Nonuniqueness of Equilibrium States: An Example.” It is easy to verify that in that example $\text{var}_{k+1}(\phi) = |a_k|$. For instance, if $a_k = -1/(k+1)^2$ there is a unique Gibbs/equilibrium state. If $a_k = -3 \log((k+1)/k)$ for $k \geq 1$ and $a_0 = -\log \sum_{j=1}^{\infty} j^{-3}$, then from Hofbauer (1977) we know that ϕ admits more than one

equilibrium state, one of them being $\delta_{11\dots}$, which cannot be a Gibbs state for any continuous ϕ .

Examples from Differentiable Dynamics

In this section we present a number of examples to which the general theory developed above does not apply directly but only after a transfer of the theory from a symbolic space to a manifold. We restrict to examples where the *results* can be transferred because those aspects of the smooth dynamics we focus on can be studied as well on a shift dynamical system that is obtained from the original one via symbolic coding. (We do not discuss the coding process itself which is sometimes far from trivial, but we focus on the application of the Gibbs and equilibrium theory.) There are alternative approaches where instead of the results the *concepts and (partly) the strategies of proofs* are transferred to the smooth dynamical systems. This has led both to an extension of the range of possible applications of the theory and to a number of refined results (because some special features of smooth systems necessarily get lost by transferring the analysis to a completely disconnected metric space).

In the following examples, T denotes a (possibly piecewise) differentiable map of a compact smooth manifold M . Points on the manifold are denoted by u and v . In all examples there is a Hölder continuous coding map $\pi : X \rightarrow M$ from a subshift of finite type X onto the manifold which respects the dynamics, that is, $T \circ \pi = \pi \circ \sigma$. This *factor map* π is “nearly” invertible in the sense that the set of points in M with more than one preimage under π has measure zero for all T -invariant measures we are interested in. Hence such measures $\tilde{\mu}$ on M correspond unambiguously to shift invariant measures $\mu = \tilde{\mu} \circ \pi^{-1}$. Similarly observables $\tilde{\phi}$ on M and $\phi = \tilde{\phi} \circ \pi$ on X are related.

Uniformly Expanding Markov Maps of the Interval

A transformation T on $M := [0, 1]$ is called a *Markov map*, if there are $0 = u_0 < u_1 < \dots < u_N = 1$

such that each restriction $T|_{(u_{i-1}, u_i]}$ is strictly monotone, C^{1+r} for some $r > 0$, and maps (u_{i-1}, u_i) onto a union of some of these N monotonicity intervals. It is called *uniformly expanding* if there is some $k \in \mathbb{N}$ such that $\lambda := \inf_x |(T^k)'(x)| > 1$. It is not difficult to verify that the symbolic coding of such a system leads to a topological Markov chain over the alphabet $A = \{1, \dots, N\}$. To simplify the discussion we assume that the transition matrix M of this topological Markov chain is irreducible and aperiodic.

Our goal is to find a T -invariant measure $\tilde{\mu}$ represented by $\mu \in \mathcal{M}_\sigma(X_M)$ which minimizes the relative entropy to Lebesgue measure on $[0, 1]$

$$h(\tilde{\mu}|m) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{a_0, \dots, a_{n-1} \in \{1, \dots, N\}} \mu[a_0 \dots a_{n-1}] \log \frac{\mu[a_0 \dots a_{n-1}]}{v_n[a_0 \dots a_{n-1}]},$$

where $v_n[a_0 \dots a_{n-1}] := |I_{a_0 \dots a_{n-1}}|$. (Recall that, without insisting on invariance, this would just be the Lebesgue measure itself.) The existence of the limit will be justified below – observe that m is not a Gibbs state as v is in Eq. (17). The argument rests on the simple observation (implied by the uniform expansion and the piecewise Hölder-continuity of T') that T has *bounded distortion*, that is, that there is a constant $C > 0$ such that for all $n \in \mathbb{N}$, $a_0 \dots a_{n-1} \in \{1, \dots, N\}^n$ and $u \in I_{a_0 \dots a_{n-1}}$ holds

$$C^{-1} \leq |I_{a_0 \dots a_{n-1}}| \cdot |(T^n)'(u)| \leq C, \text{ or, equivalently,} \\ C^{-1} \leq \frac{|I_{a_0 \dots a_{n-1}}|}{\exp(S_n \tilde{\phi}(u))} \leq C, \quad (21)$$

where $\tilde{\phi}(u) := -\log |T'(u)|$. (Observe the similarity between this property and the Gibbs property (16).) Assuming bounded distortion we have at once

$$h(\tilde{\mu}|m) = \lim_{n \rightarrow \infty} \frac{1}{n} \left(-H_n(\mu) - \sum_{k=0}^{n-1} \langle \phi \circ \sigma^k, \mu \rangle \right) \\ = -h(\mu) - \langle \phi, \mu \rangle,$$

and minimizing this relative entropy just amounts to maximizing $h(\mu) + \langle \phi, \mu \rangle$ for $\phi = -\log |T'| \circ \pi$. As the results on Gibbs distributions from section “[The Gibbs Property: A Local Characterization of Equilibrium](#)” apply, we conclude that

$$C^{-1} \leq \frac{\mu[a_0 \dots a_{n-1}]}{|I_{a_0 \dots a_{n-1}}|} \leq C$$

for some $C > 0$. So the unique T -invariant measure $\tilde{\mu}$ that minimizes the relative entropy $h(\tilde{\mu}|m)$ is equivalent to Lebesgue measure m . (The existence of an invariant probability measure equivalent to m is well known, also without invoking entropy theory. It is guaranteed by a “Folklore Theorem” (Jakobson 2002).)

Interval Maps with an Indifferent Fixed Point

The presence of just one point $x \in [0, 1]$ such that $T'(x) = 1$ dramatically changes the properties of the system. A canonical example is the map $T_\alpha: x \mapsto x(1 + 2^\alpha x^\alpha)$ if $x \in [0, 1/2]$ and $x \mapsto 2x - 1$ if $x \in [1/2, 1]$. We have $T'(0) = 1$, that is, 0 is an indifferent fixed point. For $\alpha \in [0, 1[$ this map admits an absolutely continuous invariant probability measure $d\mu(x) = h(x)dx$, where $h(x) \sim x^{-\alpha}$ when $x \rightarrow 0$ (Thaler 1980). In the physics literature, this type of map is known as the “Manneville-Pomeau” map. It was introduced as a model of transition from laminar to intermittent behavior (Pomeau and Manneville 1980). In Gaspard and Wang (1988) the authors construct a piecewise affine version of this map to study the complexity of trajectories (in the sense of subsection “[A Short Digression on Complexity](#)”). This gives rise to a countable state Markov chain. In Wang (1989) the close connection to the Fisher-Felderhof model and Hofbauer’s example (see subsection “[Non-uniqueness of Equilibrium States: An Example](#)”) was realized. We refer to Sarig (2001) for recent developments and a list of references.

Axiom A Diffeomorphisms, Anosov Diffeomorphisms, Sinai-Ruelle-Bowen Measures

The first spectacular application of the theory of Gibbs measures to differentiable dynamical

systems was Sinai's approach to Anosov diffeomorphisms via Markov partitions (Sinai 1968) that allowed one to code the dynamics of these maps into a subshift of finite type and to study their invariant measures by methods from equilibrium statistical mechanics (Sinai 1972) that had been developed previously by Dobrushin, Lanford, and Ruelle (Dobrushin 1968a, b, c, 1969; Lanford and Ruelle 1969). Not much later this approach was extended by Bowen (1970) to Smale's Axiom A diffeomorphisms (and to Axiom A flows by Bowen and Ruelle (1975)); see also Ruelle (1976). The interested reader can consult, for example, Young (2002) for a survey, and either (Bowen 2017) or (Chernov 2002) for details.

Both types of diffeomorphisms act on a smooth compact Riemannian manifold M and are characterized by the existence of a compact T -invariant *hyperbolic set* $\Lambda \subseteq M$. Their basic properties are described in detail in the contribution ► “Ergodic Theory: Basic Examples and Constructions.” Very briefly, the tangent bundle over Λ splits into two invariant subbundles – a stable one and an unstable one. Correspondingly, through each point of Λ there passes a local stable and a local unstable manifold which are both tangent to the respective subspaces of the local tangent space. The unstable derivative of T , that is, the derivative DT restricted to the unstable subbundle, is uniformly expanding. Its Jacobian determinant, denoted by $J^{(u)}$, is Hölder continuous as a function on Λ . Hence the observable $\phi^{(u)} := -\log |J^{(u)}| \circ \pi$ is Hölder continuous, and the Gibbs and equilibrium theory apply (via the symbolic coding) to the diffeomorphism T (modulo possibly a decomposition of the hyperbolic set into irreducible and aperiodic components, called basic sets, that can be modeled by topologically mixing subshifts of finite type). The main results are:

Characterization of attractors The following assertions are equivalent for a basic set $\Omega \subseteq \Lambda$:

1. Ω is an attractor, that is, there are arbitrarily small neighborhoods $U \subseteq M$ of Ω such that $TU \subset U$.

2. The union of all stable manifolds through points of Ω is a subset of M with positive volume.
3. The pressure $P_{T|_{\Omega}}(\phi^{(u)}) = 0$.

In this case the unique equilibrium and Gibbs state μ^+ of $T|_{\Omega}$ is called the *Sinai-Ruelle-Bowen (SRB) measure* of $T|_{\Omega}$. It is uniquely characterized by the identity $h_{T|_{\Omega}}(\mu^+) = -\langle \phi^{(u)}, \mu^+ \rangle$. (For all other T -invariant measures on Ω one has “ $<$ ” instead of “ $=$.”)

Further properties of SRB measures Suppose $P_{T|_{\Omega}}(\phi^{(u)}) = 0$ and let μ^+ be the SRB measure.

1. For a set of points $u \in M$ of positive volume we have:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k u) = \langle f, \mu^+ \rangle.$$

(Indeed, because of (ii) of the above characterization, this holds for almost all points of the union of the stable manifolds through points of Ω .)

2. Conditioned on unstable manifolds, μ^+ is absolutely continuous to the volume measure on unstable manifolds.

In the special case of transitive Anosov diffeomorphisms, the whole manifold is a hyperbolic set and $\Omega = M$. Because of transitivity, property (ii) from the characterization of attractors is trivially satisfied, so there is always a unique SRB measure μ^+ . As T^{-1} is an Anosov diffeomorphism as well – only the roles of stable and unstable manifolds are interchanged – T^{-1} has a unique SRB measure μ^- which is the unique equilibrium state of T^{-1} (and hence also of T) for $\phi^{(s)} := \log |J^{(s)}|$. One can show:

SRB measures for Anosov diffeomorphisms The following assertions are equivalent:

1. $\mu^+ = \mu^-$.
2. μ^+ or μ^- is absolutely continuous w.r.t the volume measure on M .
3. For each periodic point $u = T^n u \in M$, $|J(u)| = 1$, where J denotes the determinant of DT .

We remark that, similarly to the case of Markov interval maps, the unstable Jacobian of T^n at u is asymptotically equivalent to the volume of the “ n -cylinder” of the Markov partition around u . So the maximization of $h(\mu) + \langle \phi^{(u)}, \mu \rangle$ by the SRB measure μ^+ can again be interpreted as the minimization of the relative entropy of invariant measures with respect to the normalized volume, and the fact that $P(\phi^{(u)}) = 0$ in the Anosov (or more generally attractor) case means that μ^+ is as close to being absolutely continuous as it is possible for a singular measure. This is reflected by the above properties (a) and (b).

We emphasize the meaning of property (a) above: it tells us that the SRB measure μ^+ is the only *physically observable* measure. Indeed, in numerical experiments with physical models, one picks an initial point $u \in M$ “at random” (i.e., with respect to the volume or Lebesgue measure) and follows its orbit $T^k u$, $k \geq 0$.

Bowen’s Formula for the Hausdorff Dimension of Conformal Repellers

Just as nearby orbits converge towards an attractor, they diverge away from a repeller. Conformal repellers form a nice class of systems which can be coded by a subshift of finite type. The construction of their Markov partitions is much simpler than that of Anosov diffeomorphisms, see, for example, Zinsmeister (2000).

Let us recall the definition of a conformal repeller before giving a fundamental example. Given a holomorphic map $T : V \rightarrow \mathbb{C}$ where $V \subset \mathbb{C}$ is open and J a compact subset of $\mathbb{C}$, one says that (J, V, T) is a conformal repeller if

1. There exist $C > 0$, $\alpha > 1$ such that $|(T^n)'(z)| \geq C\alpha^n$ for all $z \in J$, $n \geq 1$.
2. $J = \bigcap_{n \geq 1} T^{-n}(V)$.
3. for any open set U such that $U \cap J \neq \emptyset$, there exists n such that $T^n(U \cap J) \supset J$.

From the definition it follows that $T(J) = J$ and $T^{-1}(J) = J$.

A fundamental example is the map $T: z \rightarrow z^2 + c$, $c \in \mathbb{C}$ being a parameter. It can be shown that for $|c| < \frac{1}{4}$ there exists a compact set J , called a

(hyperbolic) *Julia set*, such that $(J, \mathbb{C}, T)$ is a conformal repeller.

Conformal repellers J are in general fractal sets and one can measure their “degree of fractality” by means of their Hausdorff dimension, $\dim_H(J)$. Roughly speaking, one computes this dimension by covering the set J by balls with radius less than or equal to δ . If $N_\delta(J)$ denotes the cardinality of the smallest such covering, then we expect that

$$N_\delta(J) \delta^{-\dim_H(J)}, \quad \text{as } \delta \rightarrow 0.$$

We refer the reader to ► “Ergodic Theory: Fractal Geometry” or Falconer (2003), Pesin (1997) for a rigorous definition (based on Carathéodory’s construction) and for more information on fractal geometry.

Bowen’s formula relates $\dim_H(J)$ to the unique zero of the pressure function $\beta \mapsto P(\beta\tilde{\phi})$ where $\tilde{\phi} := -(\log |T'|)|_J$. It is not difficult to see that indeed this map has a unique zero for some positive β .

By property (1), $S_n\tilde{\phi} \leq \text{const} - n \log \alpha$, which implies (by (13)) that $\frac{d}{d\beta} P(\beta\tilde{\phi}) = \langle \tilde{\phi}, \mu_\beta \rangle \leq -\log \alpha < 0$. As $P(0)$ equals the topological entropy of J , that is, the logarithm of the largest eigenvalue of the matrix M associated to the Markov partition, $P(0)$ is strictly positive. Therefore, (recall that the pressure function is continuous) there exists a unique number $\beta_0 > 0$ such that $P(\beta_0\tilde{\phi}) = 0$.

It turns out that this unique zero is precisely $\dim_H(J)$:

Bowen’s formula The Hausdorff dimension of J is the unique solution of the equation $P(\beta\tilde{\phi}) = 0$, $\beta \in \mathbb{R}$; in particular

$$P(\dim_H(J)\tilde{\phi}) = 0.$$

This formula was proven in Ruelle (1982) for a general class of conformal repellers after the seminal paper (Bowen 1979). The main tool is a distortion estimate very similar to (21). A simple exposition can be found in Zinsmeister (2000).

Nonequilibrium Steady States and Entropy Production

SRB measures for Anosov diffeomorphisms and Axiom A attractors have been accepted recently as conceptual models for *nonequilibrium steady states* in nonequilibrium statistical mechanics. Let us point out that the word “equilibrium” is used in physics in a much more restricted sense than in ergodic theory. Only diffeomorphisms preserving the natural volume of the manifold (or a measure equivalent to the volume) would be considered as appropriate toy models of physical equilibrium situations. In the case of Anosov diffeomorphisms this is precisely the case if the “forward” and “backward” SRB measures μ^+ and μ^- coincide. Otherwise, the diffeomorphism models a situation out of equilibrium, and the difference between μ^+ and μ^- can be related to entropy production and irreversibility.

Gallavotti and Cohen (1995; Gallavotti 1996) introduced SRB measures as idealized models of nonequilibrium steady states around 1995. In order to have as firm a mathematical basis as possible they made the “*chaotic hypothesis*” that the systems they studied behave like transitive Anosov systems. Ruelle (1996) extended his approach to more general (even nonuniformly) hyperbolic dynamics; see also his reviews (Ruelle 1998, 2003) for more recent accounts discussing also a number of related problems; see by Rondoni and Mejía-Monasterio (2007), too. The importance of the Gibbs property of SRB measures for the discussion of entropy production was also highlighted in Jiang et al. (2000), where it is shown that for transitive Anosov diffeomorphisms the relative entropy $h(\mu^+|\mu^-)$ equals the average entropy production rate $\langle \log|J|, \mu^+ \rangle$ of μ^+ where J denotes again the Jacobian determinant of the diffeomorphism. In particular, the entropy production rate is zero if, and only if, $h(\mu^+|\mu^-) = 0$, that is, using coding and (19), if, and only if, $\mu^+ = \mu^-$. According to subsection “*Axiom A Diffeomorphisms, Anosov Diffeomorphisms, Sinai-Ruelle-Bowen Measures*,” this is also equivalent to μ^+ or μ^- being absolutely continuous with respect to the volume measure.

Some Ongoing Developments and Future Directions

As we saw, many dynamical systems with uniform hyperbolic structure (e.g., Anosov maps, axiom A diffeomorphisms) can be modeled by subshifts of finite type over a finite alphabet. We already mentioned in subsection “*Interval Maps with an Indifferent Fixed Point*” the typical example of a map of the interval with an indifferent fixed point, whose symbolic model is still a subshift of finite type, but with a countable alphabet. The thermodynamic formalism for such systems is by now well developed (Fiebig et al. 2002; Gurevich and Savchenko 1998; Sarig 1999, 2001, 2003) and used, for example, for multidimensional piecewise expanding maps (Buzzi and Sarig 2003). An active line of research is related to systems admitting representations by symbolic models called “towers” constructed by using “inducing schemes.” The fundamental example is the class of one-dimensional unimodal maps satisfying the “Collet-Eckmann condition.” A first attempt to develop thermodynamic formalism for such systems was made in Bruin and Keller (1998) where existence and uniqueness of equilibrium measures for the potential function $\tilde{\phi}_\beta(u) = -\beta \log |T'(u)|$ with β close to 1 was established. Very recently, new developments in this direction appeared, see, for example, (Bruin and Todd 2008, 2009; Pesin and Senti 2008).

A largely open field of research concerns a new branch of nonequilibrium statistical mechanics, the so-called chaotic scattering theory, namely the analysis of chaotic systems with various openings or holes in phase space, and the corresponding repellers on which interesting invariant measures exist. We refer the reader to Chernov (2002) for a brief account and references to the physics literature. The existence of (generalized) steady states on repellers and the so-called escape rate formula have been observed numerically in a number of models. So far, little has been proven mathematically, except for Anosov diffeomorphisms with special holes (Chernov 2002) and for certain nonuniformly hyperbolic systems (Bruin et al. 2010).

Bibliography

- Baladi V (2000) Positive transfer operators and decay of correlations. Advanced series in nonlinear dynamics, vol 16. World Scientific, Singapore
- Bowen R (1970) Markov partitions for Axiom A diffeomorphisms. *Am J Math* 92:725–747
- Bowen R (1974a/1975) Some systems with unique equilibrium states. *Math Syst Theory* 8:193–202
- Bowen R (1974b/75) Bernoulli equilibrium states for Axiom A diffeomorphisms. *Math Syst Theory* 8: 289–294
- Bowen R (1979) Hausdorff dimension of quasicircles. *Inst Hautes Études Sci Publ Math* 50:11–25
- Bowen R (2017) Equilibrium states and the ergodic theory of Anosov diffeomorphisms. Lecture notes in mathematics, vol 470, 2nd edn (1st edn 1975). Springer, Berlin
- Bowen R, Ruelle D (1975) The ergodic theory of Axiom A flows. *Invent Math* 29:181–202
- Bruđno AA (1983) Entropy and the complexity of the trajectories of a dynamical system. *Trans Mosc Math Soc* 2:127–151
- Bruin H, Keller G (1998) Equilibrium states for S -unimodal maps. *Ergod Theory Dyn Syst* 18(4): 765–789
- Bruin H, Todd M (2008) Equilibrium states for potentials with $\sup \varphi - \inf \varphi < h_{\text{top}}(f)$. *Commun Math Phys* 283(3):579–611. <https://doi.org/10.1007/s00220-0-008-0596-0>
- Bruin H, Todd M (2009) Equilibrium states for interval maps: the potential $-t \log |Df|$. *Ann Sci Éc Norm Sup* (4) 42(4): 559–600. <https://doi.org/10.24033/asens.2103>
- Bruin H, Demers M, Melbourne I (2010) Existence and convergence properties of physical measures for certain dynamical systems with holes. *Ergod Theory Dyn Syst* 30(3):687–728. <https://doi.org/10.1017/S0143385709000200>
- Buzzi J, Sarig O (2003) Uniqueness of equilibrium measures for countable Markov shifts and multi-dimensional piecewise expanding maps. *Ergod Theory Dyn Syst* 23(5):1383–1400
- Chaitin GJ (1987) Information, randomness & incompleteness. Papers on algorithmic information theory. World Scientific series in computer science, vol 8. World Scientific, Singapore
- Chernov N (2002) Invariant measures for hyperbolic dynamical systems. In: Handbook of dynamical systems, vol 1A. North-Holland, pp 321–407
- Coelho Z, Parry W (1990) Central limit asymptotics for shifts of finite type. *Israel J Math* 69(2):235–249
- Dobrushin RL (1968a) The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory Probab Appl* 13: 197–224
- Dobrushin RL (1968b) Gibbsian random fields for lattice systems with pairwise interactions. *Funct Anal Appl* 2: 292–301
- Dobrushin RL (1968c) The problem of uniqueness of a Gibbsian random field and the problem of phase transitions. *Funct Anal Appl* 2:302–312
- Dobrushin RL (1969) Gibbsian random fields. The general case. *Funct Anal Appl* 3:22–28
- Eizenberg A, Kifer Y, Weiss B (1994) Large deviations for $\mathbb{Z}^d$ -actions. *Commun Math Phys* 164(3):433–454
- Falconer K (2003) Fractal geometry. Mathematical foundations and applications, 2nd edn. Wiley, San Francisco
- Fiebig D, Fiebig U-R, Yuri M (2002) Pressure and equilibrium states for countable state Markov shifts. *Israel J Math* 131:221–257
- Fisher ME (1967) The theory of condensation and the critical point. *Physics* 3:255–283
- Fisher ME, Felderhof BU (1970) Phase transition in one-dimensional clusterinteraction fluids: IA. Thermodynamics, IB. Critical behavior. II. Simple logarithmic model. *Ann Phys* 58:177–280
- Gallavotti G (1996) Chaotic hypothesis: Onsager reciprocity and fluctuation-dissipation theorem. *J Stat Phys* 84: 899–925
- Gallavotti G, Cohen EGD (1995) Dynamical ensembles in stationary states. *J Stat Phys* 80:931–970
- Gaspard P, Wang X-J (1988) Sporadicity: between periodic and chaotic dynamical behaviors. *Proc Natl Acad Sci USA* 85:4591–4595
- Georgii H-O (1988) Gibbs measures and phase transitions. In: de Gruyter studies in mathematics. de Gruyter, Berlin, p 9
- Gurevich BM, Savchenko SV (1998) Thermodynamic formalism for symbolic Markov chains with a countable number of states. *Russ Math Surv* 53(2):245–344
- Hofbauer F (1977) Examples for the nonuniqueness of the equilibrium state. *Trans Am Math Soc* 228:223–241
- Israel R (1979) Convexity in the theory of lattice gases. Princeton series in physics. Princeton University Press
- Jakobson M, Świątek (2002) One-dimensional maps. In: Handbook of dynamical systems, vol 1A. North-Holland, Amsterdam, pp 321–407
- Jaynes ET (1989) Papers on probability, statistics and statistical physics. Kluwer, Dordrecht
- Jiang D, Qian M, Qian M-P (2000) Entropy production and information gain in Axiom A systems. *Commun Math Phys* 214:389–409
- Katok A (2007) Fifty years of entropy in dynamics: 1958–2007. *J Mod Dyn* 1(4):545–596
- Katok A, Hasselblatt B (1995) Introduction to the modern theory of dynamical systems. Encyclopaedia of mathematics and its applications, vol 54. Cambridge University Press, Cambridge
- Keller G (1998) Equilibrium states in ergodic theory. In: London Mathematical Society student texts, vol 42. Cambridge University Press, Cambridge
- Kifer Y (1990) Large deviations in dynamical systems and stochastic processes. *Trans Am Math Soc* 321:505–524
- Kolmogorov AN (1983) Combinatorial foundations of information theory and the calculus of probabilities. *Uspekhi Mat Nauk* 38:27–36

- Lanford OE, Ruelle D (1969) Observables at infinity and states with short range correlations in statistical mechanics. *Commun Math Phys* 13:194–215
- Lewis JT, Pfister C-E (1995) Thermodynamic probability theory: some aspects of large deviations. *Russ Math Surv* 50:279–317
- Lind D, Marcus B (1995) An introduction to symbolic dynamics and coding. Cambridge University Press, Cambridge
- Misiurewicz M (1976) A short proof of the variational principle for a $\mathbb{Z}_+^N$ action on a compact space. In: International conference on dynamical systems in mathematical physics (Rennes, 1975), *Astérisque*, No. 40, Soc Math France, pp 147–157
- Moulin-Ollagnier J (1985) Ergodic theory and statistical mechanics. In: *Lecture notes in mathematics*, vol 1115. Springer, Berlin
- Pesin Y (1997) Dimension theory in dynamical systems. Contemporary views and applications. University of Chicago Press, Chicago
- Pesin Y, Senti S (2008) Equilibrium measures for maps with inducing schemes. *J Mod Dyn* 2(3):397–430. <https://doi.org/10.3934/jmd.2008.2.397>
- Pesin Y, Weiss (1997) The multifractal analysis of Gibbs measures: motivation, mathematical foundation, and examples. *Chaos* 7(1):89–106
- Pollicott M (2000) Rates of mixing for potentials of summable variation. *Trans Am Math Soc* 352(2):843–853
- Pomeau Y, Manneville P (1980) Intermittent transition to turbulence in dissipative dynamical systems. *Commun Math Phys* 74(2):189–197
- Rondoni L, Mejía-Monasterio C (2007) Fluctuations in nonequilibrium statistical mechanics: models, mathematical theory, physical mechanisms. *Nonlinearity* 20(10):R1–R37
- Ruelle D (1968) Statistical mechanics of a one-dimensional lattice gas. *Commun Math Phys* 9:267–278
- Ruelle D (1973) Statistical mechanics on a compact set with $\mathbb{Z}^v$ action satisfying expansiveness and specification. *Trans Am Math Soc* 185:237–251
- Ruelle D (1976) A measure associated with Axiom A attractors. *Am J Math* 98:619–654
- Ruelle D (1982) Repellers for real analytic maps. *Ergod Theory Dyn Syst* 2(1):99–107
- Ruelle D (1996) Positivity of entropy production in nonequilibrium statistical mechanics. *J Stat Phys* 85:1–23
- Ruelle D (1998) Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics. *J Stat Phys* 95:393–468
- Ruelle D (2003) Extending the definition of entropy to nonequilibrium steady states. *Proc Nat Acad Sci USA* 100(6):3054–3058
- Ruelle D (2004) Thermodynamic formalism: the mathematical structures of equilibrium statistical mechanics, Cambridge Mathematical Library, 2nd edn. Cambridge University Press, Cambridge
- Sarig O (1999) Thermodynamic formalism for countable Markov shifts. *Ergod Theory Dyn Syst* 19(6):1565–1593
- Sarig O (2001) Phase transitions for countable Markov shifts. *Commun Math Phys* 217(3):555–577
- Sarig O (2003) Existence of Gibbs measures for countable Markov shifts. *Proc Am Math Soc* 131(6):1751–1758
- Seneta E (2006) Non-negative matrices and Markov chains. In: *Springer series in statistics*. Springer
- Sinai JG (1968) Markov partitions and C-diffeomorphisms. *Funct Anal Appl* 2:61–82
- Sinai JG (1972) Gibbs measures in ergodic theory. *Russ Math Surv* 27(4):21–69
- Thaler M (1980) Estimates of the invariant densities of endomorphisms with indifferent fixed points. *Israel J Math* 37(4):303–314
- Walters P (1975) Ruelle's operator theorem and g -measures. *Trans Am Math Soc* 214:375–387
- Walters P (1992) Differentiability properties of the pressure of a continuous transformation on a compact metric space. *J Lond Math Soc* (2) 46(3):471–481
- Wang X-J (1989) Statistical physics of temporal intermittency. *Phys Rev A* 40(11):6647–6661
- Young L-S (1990) Large deviations in dynamical systems. *Trans Am Math Soc* 318:525–543
- Young L-S (2002) What are SRB measures, and which dynamical systems have them? Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays. *J Stat Phys* 108(5–6):733–754
- Zinsmeister M (2000) Thermodynamic formalism and holomorphic dynamical systems. *SMF/AMS texts and monographs*, vol 2. American Mathematical Society

Parallels Between Topological Dynamics and Ergodic Theory

Wen Huang, Song Shao and Xiangdong Ye
CAS Wu Wen-Tsun Key Laboratory of
Mathematics, and Department of Mathematics,
University of Science and Technology of China,
Hefei, Anhui, China

Article Outline

Glossary
Definition of the Subject
Introduction and History
Recurrence and Other Dynamical Properties
Entropy Theory
Structure Theorems and Multiple Ergodic
Averages
Further Directions
References

Glossary

Ergodicity A measure-preserving system is ergodic if it is essentially indecomposable, in the sense that given any invariant measurable set, either the set or its complement has measure 0.

Measure-preserving transformation A map from a measure space to itself such that for each measurable subset of the space, it has the same measure as its inverse image under the map. Such a measure-preserving map on a measure space is called a measure-preserving system.

Measure-theoretic and topological entropy A nonnegative (possibly infinite) real number which describes the complexity of a measure-preserving transformation and topological dynamics.

(Multiple) Ergodic average Let $(X, \mathcal{X}, \mu, T)$ be a measure-preserving system and f be a real-

valued (or complex-valued) function on X . Then $A_N(f, x) = \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)$ is called the ergodic average. The multiple ergodic averages (or called “nonconventional averages”) are the following ones

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T_1^{p_1(n)} x) f_2(T_2^{p_2(n)} x) \dots f_d(T_d^{p_d(n)} x),$$

where $T_1, T_2, \dots, T_d$ are invertible and act on a probability space $(X, \mathcal{X}, \mu)$, $f_1, \dots, f_d$ are functions on X , and $p_1, \dots, p_d$ are integral polynomials.

Recurrent point A point is recurrent if it is in its own future.

Topological dynamics The study of the asymptotic behaviors of a homeomorphism from a topological space to itself. Such a self-homeomorphism of a topological space is called a topological dynamical system.

Transitive point A point is a transitive point when every point is in its future.

Transitivity and minimality A system is transitive when it contains at least one transitive point. A system is minimal when every point is a transitive point.

Definition of the Subject

By topological dynamics, it is “the study of transformation groups with respect to those topological properties whose prototype occurred in classical dynamics” (Gottschalk and Hedlund 1955). If in this definition the adjective “topological” is replaced by “measure-theoretic,” then one obtains a description of measurable dynamics, also called ergodic theory.

In this entry we try to present theorems illustrating the analogy between topological and measurable dynamics, and theorems showing drastic contrast between them.

Introduction and History

In its broadest interpretation a dynamical system is the study of the qualitative properties of actions of groups on spaces. The space has some structure (topological, measurable, smooth manifold, etc.) and the action preserves the structure (so acts as a continuous map or a homeomorphism, a measure preserving transformation, a differential map, etc.). The dynamical systems are then divided into branches including topological dynamics, ergodic theory, differentiable dynamical systems, etc., according to the spaces and the actions. In this entry we mainly present some aspects of the brother branches: topological dynamics and ergodic theory.

The terminology used in both branches is almost the same. For example, one talks about transitivity or minimality, topological weak/mild/strong mixing, distality, rigidity, disjointness in topological dynamics. At the same time, one speaks ergodicity, measurable weak/mild/strong mixing, distality, rigidity, disjointness in ergodic theory. These have led to the formulation of analogous theorems. However, they have up until now remained analogies; the proofs involved being entirely different and neither directly deducible from the other, for example, the proofs of the structure theorems in both theories. This, of course, is not surprising since the methods used in one are topological and in the other measure theoretic. We remark that some topological statements are proved first by using the measure theoretical methods and then one finds the topological arguments, for example, the fact that the a minimal system with a trivial regionally proximal relation of order d is a d -step pro-nilpotent system (Host et al. 2010; Cai and Shao 2019; Candela 2017a, b; Gutman et al. 2019a, b, 2020). Moreover, some topological statements are proved by using the measure theoretical methods, and one has no topological proofs until now, for example, the statement that positive entropy implies the existence of a proper asymptotic pair (Blanchard et al. 2002b).

This kind research arises from the study of the ordinary differential equation. We will skip the description and briefly mention the histories of topological dynamics and ergodic theory.

Topological dynamics owes its origin to the classic work of Poincaré and Birkhoff (1927). It was Poincaré who first formulated and solved problems of dynamics as problems in topology. An autonomous system of ordinary differential equations satisfying conditions that ensure uniqueness and extendibility of solutions determines a flow, that is, a one-parameter transformation group. Birkhoff was the first to note that many notions and results obtained in the theory of autonomous differential equations can be extended to the case of flows in abstract spaces. Introducing the very important concepts of minimal sets and of recurrent and central motions, he laid the foundation of the general theory of flows or, it is usually called, the topological theory of dynamical systems. This theory was further developed by lots of mathematicians (Nemytskiĭ 1949; Nemytskiĭ and Stepanov 1949). At the end of 1940s Gottschalk and Hedlund and many others proposed a broad and very natural generalization of classical dynamical systems, namely, the notion of a topological transformation group (Nemytskiĭ and Stepanov 1949; Gottschalk and Hedlund 1955; Ellis 1969). The part of the theory of transformation groups devoted to the study of those notions whose prototype occurred previously in classical dynamics, such as various types of motions, minimal and limit set, recurrence, and stability, is called topological dynamics.

Ergodic theory is not one of the classical mathematical disciplines and its name, in contrast to, for example, measure theory and number theory, does not indicate its subject. However, its origin can be described quite precisely. Ergodic theory has its roots in Maxwell's and Boltzmann's kinetic theory of gases. The first important result was found by Poincaré by the end of the nineteenth century, which now is called Poincaré recurrence theorem. Ergodic theory was born as a mathematical theory around 1930 by the ground breaking works of von Neumann and Birkhoff. Two other major contributions should be mentioned. One is the introduction of the notion of entropy by Kolmogorov and Sinai at the end of the 1950s. Another is Ornstein's theorem saying that the entropy is a complete invariant for Bernoulli

shifts. In the 1970s, Furstenberg showed how to translate questions in combinatorial number theory into ergodic theory. This inspired a new line of research, which ultimately led to stunning recent results in combinatorial number theory.

For simplicity in this article we only consider $\mathbb{Z}$ -actions. Thus, by a *topological dynamics* (t.d.s. for short) we mean a pair (X, T) , where X is a compact metrizable space with metric ρ and $T : X \rightarrow X$ is a homeomorphism. By a *measure-preserving system* (m.p.s. for short) we mean a quadruple $(X, \mathcal{X}, \mu, T)$, where $(X, \mathcal{X}, \mu)$ is a Borel probability space and $T, T^{-1} : X \rightarrow X$ are both measurable and measure preserving, that is, $T^{-1}\mathcal{X} = \mathcal{X} = T\mathcal{X}$ and $\mu(A) = \mu(T^{-1}A)$ for each $A \in \mathcal{X}$.

There are two types of problems in topological dynamics and ergodic theory. The first type can be viewed as the internal problems which concern with understanding the homeomorphisms or measure preserving transformations and trying to decide when two of them are conjugate or isomorphic. The second type is the applications of the theories to other branches of mathematics or in physics. For the internal problems, the usual way is to look for conjugacy or isomorphic invariants. Such invariants could be a property (e.g., minimality, ergodicity), or is an assignment of some object (e.g., a number like entropy, a group or a structure). For more details, see (Walters 1982).

Let us now introduce the notions of conjugacy and isomorphism. Let (X, T) and (Y, S) be two t.d.s. A continuous map $\pi : X \rightarrow Y$ is called a *homomorphism* or *factor map* between (X, T) and (Y, S) if it is onto and $\pi \circ T = S \circ \pi$. In this case we say (X, T) is an *extension* of (Y, S) or (Y, S) is a *factor* of (X, T) . When π is a homeomorphism, we then say that (X, T) and (Y, S) are *conjugate*, so in this case we review the two systems are the same.

To define the same notion in ergodic theory, we keep in mind that a null set is negligible. Suppose $(X_i, \mathcal{X}_i, \mu_i)$ is a probability space and $T_i : X_i \rightarrow X_i$ is measure preserving, $i = 1, 2$. We say that T_2 is a factor of T_1 or T_1 is an extension of T_2 if there exists $M_i \in \mathcal{X}_i$ with $\mu_i(M_i) = 1$ and $T_i(M_i) \subset M_i$ ($i = 1, 2$) and there exists a measure preserving transformation $\phi : M_1 \rightarrow M_2$ with $\phi T_1(x) = T_2 \phi(x)$ for any $x \in M_1$. If ϕ is 1-1 and ϕ^{-1} is also measure preserving, then we say ϕ is an *isomorphism*.

Let (X, T) be a t.d.s. By Krylov-Bogolioubov theorem, there is at least one invariant probability measure μ on the Borel σ -algebra $\mathcal{B}_X$. In such a way, $(X, \mathcal{B}_X, T, \mu)$ can be viewed as an m.p.s. which also explains why some topological results can be proved via ergodic theory.

On the other hand, for any m.p.s. $(X, \mathcal{X}, \mu, T)$ one can find some t.d.s. $(\widehat{X}, \widehat{T})$ and an invariant measure $\widehat{\mu}$ on the Borel σ -algebra $\mathcal{B}_{\widehat{X}}$ such that $(X, \mathcal{X}, \mu, T)$ is isomorphic to $(\widehat{X}, \mathcal{B}_{\widehat{X}}, \widehat{\mu}, \widehat{T})$ (Furstenberg 1981). $(\widehat{X}, \widehat{T})$ is called a topological model of $(X, \mathcal{X}, \mu, T)$. This makes the possibility to use topological methods in the study of ergodic theory.

We try our best to exhibit the similarity or parallels between two theories, from the notations to the results. There is a very nice survey by Glasner and Weiss (2005) on the same subject, here we will review the subject in some new angles and present some new progress.

Due to our personnel interest, knowledge, and the space available we can only choose some aspects to do so. Thus, the materials we choose are restricted, and are the ones we are familiar with. In the first part of the entry we will discuss dynamical properties where most of results are classical. In the second part we focus on the study of entropy, particularly on the so-called local entropy theory which is relatively new. In the third part we will consider the structure theorems and their applications to the combinatorial number theory which are the hot points of current research and developing very fast. We hope in such a way we could cover certain old and new results, and some materials from the internal problems to the applications of the theories.

The authors would like to thank B. Kra, B. Host, and M. Lemańczyk for valuable comments which improved the writing of the survey significantly.

Recurrence and Other Dynamical Properties

Given a dynamical system, we are interested in those points that can be observed repeatedly. This leads to the notion of *recurrence*. The basic fact is

that any t.d.s. has a recurrent point and for any measurable set with positive measure there are points returning to the original set infinitely many times. We are particularly interested in the systems that are “non-decomposable” in some sense. Such systems are minimal or transitive systems in topological dynamics and ergodic systems in ergodic theory. Below we will see that when studying dynamical properties, some interesting subsets of $\mathbb{N}$, $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$ or $\mathbb{Z}$ will be involved. This indicates that a dynamical system has a close relation with the combinatorial number theory. We will find many evidence in the sequel.

Since many statements of the paper are better stated using the notion of a family, we now give the definition. See Akin (1997), Furstenberg (1967, 1981) for more details.

We define family on $\mathbb{Z}_+$, and it works for $\mathbb{N}$, $\mathbb{Z}$, etc. A collection $\mathcal{F}$ of subsets of $\mathbb{Z}_+$ is a *family* if it is hereditary upward, that is, $F_1 \subset F_2$ and $F_1 \in \mathcal{F}$ imply $F_2 \in \mathcal{F}$. A family $\mathcal{F}$ is called *proper* if it is neither empty nor the entire power set of $\mathbb{Z}_+$, or, equivalently if $\mathbb{Z}_+ \in \mathcal{F}$ and $\emptyset \notin \mathcal{F}$. Any nonempty collection $\mathcal{A}$ of subsets of $\mathbb{Z}_+$ generates a family

$$\mathcal{F}(\mathcal{A}) := \{F \subset \mathbb{Z}_+ : F \supset A \text{ for some } A \in \mathcal{A}\}.$$

For a family $\mathcal{F}$ its *dual* is the family

$$\begin{aligned} \mathcal{F}^* &= \{F \subset \mathbb{Z}_+ : F \cap F' \neq \emptyset \text{ for all } F' \in \mathcal{F}\} \\ &= \{F \subset \mathbb{Z}_+ : \mathbb{Z}_+ \setminus F \notin \mathcal{F}\} \end{aligned}$$

It is not hard to see that if $\mathcal{F}$ is a family, then

$$(\mathcal{F}^*)^* = \mathcal{F}.$$

If a family $\mathcal{F}$ is closed under finite intersections and is proper, then it is called a *filter*. A family $\mathcal{F}$ has the *Ramsey property* if $A = A_1 \cup A_2 \in \mathcal{F}$ implies that $A_1 \in \mathcal{F}$ or $A_2 \in \mathcal{F}$. It is well known that a proper family has the Ramsey property if and only if its dual $\mathcal{F}^*$ is a filter (Furstenberg 1981).

Minimality Versus Ergodicity

Minimality and Transitivity

The first recurrence property we will discuss is the minimality. Recall that we only consider homeomorphisms. A t.d.s. is *minimal* if the orbit of any point is dense in the space, where the *orbit* of x is the set $\{T^n x : n \in \mathbb{Z}_+\}$. Given a t.d.s. (X, T) , we say (Y, T) is a *subsystem* if Y is a closed subset of X and is invariant under T , that is, $TY = Y$. The minimality is equivalent to the statement that there is no non-empty proper invariant closed subset. We say x is a *periodic point* if there is $n \in \mathbb{N}$ such that $T^n(x) = x$. If x is a periodic point, then it is easy to see that the orbit of x is minimal. Basic examples of minimal t.d.s can be found in books (Auslander 1988; Ellis 1969; Kurka 2003; de Vries 1993), etc.

Now we consider a weaker notion than minimality, namely *transitivity*. A t.d.s. (X, T) is *transitive* (for a survey, see Kolyada and Snoha (1997)) if there is some point x with dense orbit. We have

Theorem 2.1 *Let (X, T) be a t.d.s. Then the following statements are equivalent:*

1. (X, T) is transitive.
2. For each non-empty open set U of X , $\bigcup_{n=1}^{\infty} T^{-n}U$ is dense in X .
3. For every non-empty open sets U, V of X , there exists $n > 0$ with $U \cap T^{-n}V \neq \emptyset$.

Let (X, T) be a t.d.s. $x \in X$ is a (*positively*) *recurrence point* if there is a sequence $n_i \rightarrow +\infty$ such that $T^{n_i}x \rightarrow x$, $i \rightarrow \infty$. So, if x is recurrent, then we can observe its motion near x infinitely many times. There are several ways to show the following basic fact.

Theorem 2.2 (Birkhoff recurrence theorem). *Any t.d.s. has a recurrence point.*

This theorem has an important generalization, namely the multiple topological recurrence theorem (Furstenberg 1981). We mention that it is equivalent to the well-known van der Waerden's

theorem (van der Waerden 1927; Furstenberg 1981).

When x is recurrent, the restriction of T to its orbit closure is onto and thus this orbit closure is a transitive subsystem. A point is a *transitive point* if the orbit closure is dense in X . We say x is a *minimal point* if the orbit closure of x is a minimal subset. We have

Theorem 2.3 *Let (X, T) be a transitive t.d.s. Then the set of transitive points is a G_δ subset of X . If (X, T) is transitive and not minimal, then the set of non-transitive points is dense in X .*

We remark that there exist transitive t.d.s. whose non-transitive points are periodic points (Downarowicz and Ye 2002).

The following well-known fact due to Furstenberg (1981) tells us how recurrence is related to an IP-set. Let $\{p_i\}_{i=1}^\infty$ be a sequence in $\mathbb{N}$. Define

$$FS(\{p_i\}) = \left\{ \sum_{j=1}^k p_{i_j} : 1 \leq i_1 < \dots < i_k, k \in \mathbb{N} \right\}.$$

A subset $F \subset \mathbb{N}$ is called an *IP-set* if it contains some $FS(\{p_i\}_{i=1}^\infty)$. Denote the family of all IP-sets by $\mathcal{F}_{ip}$. The well-known Hindman's theorem (Hindman 1974) states that if $N_1 \cup \dots \cup N_k$ is a partition of $\mathbb{N}$, then one of the cell contains an IP-set. It is equivalent to say that $\mathcal{F}_{ip}$ has the Ramsey property, and thus the dual family $\mathcal{F}_{ip}^*$ is a filter. Any set in $\mathcal{F}_{ip}^*$ is called an *IP*-set*.

For a t.d.s. (X, T) , $x \in X$ and $U \subset X$ let

$$N(x, U) = \{n \in \mathbb{Z}_+ : T^n x \in U\}.$$

Theorem 2.4 *Let (X, T) be a t.d.s. Then x is a recurrent point if and only if for each neighborhood U of x , $N(x, U)$ is an IP set.*

We remark that for each IP-set R , there is some t.d.s. (X, T) such that there is a recurrent point $x \in X$ and an open neighborhood U with $R \cup \{0\} \supset N(x, U)$ (Furstenberg 1981).

The minimality property is related the syndetic subsets. A subset S of $\mathbb{Z}_+$ is *syndetic* if it has a bound on the size of the gaps, that is, there is $N \in \mathbb{N}$ such that $\{i, i+1, \dots, i+N\} \cap S \neq \emptyset$

for every $i \in \mathbb{Z}_+$; S is *thick* if it contains arbitrarily long runs of positive integers, that is, for every $n \in \mathbb{N}$ there exists some $a_n \in \mathbb{Z}_+$ such that $\{a_n, a_n+1, \dots, a_n+n\} \subset S$. It is easy to show that the dual family of syndetic subsets is the family of thick subsets, and vice versa. The following fact was due to Gottschalk (n.d.) (see also Gottschalk and Hedlund (1955)).

Theorem 2.5 *For any t.d.s. (X, T) , $x \in X$ is a minimal point if and only if $N(x, U)$ is syndetic for each neighborhood U of x .*

The fact that any t.d.s. has a minimal subset follows from the Zorn's lemma, and one can find a constructive proof in Weiss (2000).

The minimality of some systems is important in the applications. For example the minimality of $(N_d(X), \langle T \times T \times \dots \times T, T \times T^2 \times \dots \times T^d \rangle)$ has an application to the simple proof of van der Waerden's theorem (Glasner 2003), where $\langle T \times T \times \dots \times T, T \times T^2 \times \dots \times T^d \rangle$ be the group generated by $T \times T \times \dots \times T$, $T \times T^2 \times \dots \times T^d$ and $N_d(X)$ is the orbit closure of $(x, \dots, x)$ under the two actions (see section “Structure Theorems and Multiple Ergodic Averages”). The minimality of $(x, \dots, x)$ under the face group actions (see section “Structure Theorems and Multiple Ergodic Averages”) has an application to the proof of the structure theorem of a minimal t.d.s. (Shao and Ye 2012).

Related to recurrence we may define a *recurrence set*. $A \subset \mathbb{N}$ is a recurrence set if for any t.d.s. (X, T) there is a point $x \in X$ and $\{n_i\} \subset A$, $i \rightarrow \infty$ such that $T^{n_i} x \rightarrow x$. Equivalently, a set A is recurrent if and only if for any minimal t.d.s. (X, T) and non-empty open set U there is some $n \in A$ such that $U \cap T^{-n} U \neq \emptyset$. We have the following fact (Furstenberg 1981).

Theorem 2.6 *A is a recurrence set if and only if for any syndetic set S , $A \cap (S - S) \neq \emptyset$.*

A deep open question related to recurrence is that: If A is a recurrence set for rotation on a torus of arbitrary dimension, is it a recurrence set? For the related research see (Weiss 2000; Ellis and Keynes 1972; Følner 1954; Katznelson 2001; Huang and Ye 2012; Huang et al. 2016; Glasner 1998; Host et al. 2016).

Ergodicity

Here we will discuss the counterpart of minimality or transitivity in ergodic theory, namely ergodicity.

We say an m.p.s. $(X, \mathcal{X}, \mu, T)$ or μ is *ergodic* if $B \in \mathcal{X}$ with $T^{-1}B = B$, then $\mu(B) = 0$ or $\mu(X \setminus B) = 0$. For example, the irrational rotation on the circle is ergodic with respect to the Lebesgue measure; the Gauss transformation is ergodic with respect to the Gauss measure (recall that the Gauss transformation $T: [0, 1) \rightarrow [0, 1)$ is given by $T(x) = 0$ if $x = 0$, and $T(x) = \{1/x\}$ if $x \neq 0$, where $\{y\}$ denotes the fractional part of a positive real number y). And the Gauss measure μ is given by $\mu(A) = \frac{1}{\log 2} \int_A \frac{1}{1+x} dx$ for any Borel set $A \subset [0, 1)$. Basic examples and constructions of ergodic m.p.s. can be found in the survey paper (Nicol and Petersen 2012), or books (Billingsley 1965; Cornfeld et al. 1982; Katok and Hasselblatt 1995). Those examples include translations on the torus, symbolic dynamical systems, interval exchange transformations, Billiards, Gauss systems, Hamiltonian systems, horocycle flows, etc. The ergodicity of some systems is important in the applications. For example, the ergodicity of the *Gauss transformation* gives precise statistical information about the continued fraction digits of almost every real number. The ergodicity of the Liouville system implies the Chowla conjecture (Frantzikinakis 2017).

The ergodicity of an m.p.s. has many equivalent statements. We list a few of them.

Theorem 2.7 *Let $(X, \mathcal{X}, \mu, T)$ be a m.p.s. Then the following statements are equivalent:*

1. T is ergodic.
2. For $A \in \mathcal{X}$ with $\mu(A) > 0$, $\mu(\bigcup_{n=1}^{\infty} T^{-n}A) = 1$.
3. For every $A, B \in \mathcal{X}$ with $\mu(A)\mu(B) > 0$ there exists $n > 0$ with $\mu(A \cap T^{-n}B) > 0$.
4. For every $A, B \in \mathcal{X}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(A \cap T^{-i}B) = \mu(A)\mu(B).$$

Similar to topological case we have the following basic fact related to recurrence.

Theorem 2.8 (*Poincaré recurrence theorem*). *Let $(X, \mathcal{X}, \mu, T)$ be an m.p.s. Then for any $A \in \mathcal{X}$ with $\mu(A) > 0$, there is $n \in \mathbb{N}$ such that $\mu(A \cap T^{-n}A) > 0$.*

The above theorem also has an important generalization, namely the multiple ergodic recurrence theorem (Furstenberg 1981). The generalized form and its relation to the well-known Szemerédi's theorem will be discussed later.

This theorem allows us to define *Poincaré sequences*. A subset $A \subset \mathbb{N}$ is a Poincaré sequence if for any m.p.s. $(X, \mathcal{X}, \mu, T)$ and for any $B \in \mathcal{X}$ with $\mu(B) > 0$ there is $n \in A$ such that $\mu(B \cap T^{-n}B) > 0$. The *upper Banach density* of a subset S of $\mathbb{Z}$ or $\mathbb{Z}_+$ is defined by

$$BD^*(S) = \limsup_{|I| \rightarrow \infty} \frac{|S \cap I|}{|I|},$$

where I ranges over intervals of $\mathbb{Z}$ or $\mathbb{Z}_+$. We have Furstenberg (1981).

Theorem 2.9 *A is a Poincaré sequence if and only if for any subset S of $\mathbb{N}$ with positive upper Banach density, $A \cap (S - S)$ contains nonzero integers.*

Thus, a Poincaré sequence is a recurrence set, and the converse does not hold (Frantzikinakis and McCutcheon 2012). Examples of a Poincaré sequence include IP-sets, $p(\mathbb{Z})$, where p is a non-constant polynomial with $p(\mathbb{Z}) \subset \mathbb{Z}$ and $p(0) = 0$, see for example (Furstenberg 1981).

Recurrence sets and Poincaré sequences have higher order generalizations, we refer to (Frantzikinakis and McCutcheon 2012; Furstenberg and Weiss 1978; Huang et al. 2016) for a systematic study of the subject and the references therein.

Ergodic Decomposition

Let $(X, \mathcal{X}, \mu, T)$ be an m.p.s. Then μ can be disintegrated as $\mu = \int_X \mu_x d\mu(x)$, where μ_x are ergodic measures. This disintegration is known as *the ergodic decomposition* of μ . The disintegration is characterized by the following properties: a) for every $f \in L^1(X, \mathcal{X}, \mu)$, $f \in L^1(X, \mathcal{X}, \mu_x)$ for μ -a.e. $x \in X$,

and the map such that $x \mapsto \int f d\mu_x$ is in $L^1(X, \mathcal{X}, \mu)$; b) for every $f \in L^1(X, \mathcal{X}, \mu)$, $\mathbb{E}_\mu(f|I(T)) \times (x) = \int f d\mu_x$ for μ -a.e. $x \in X$, where $I(T) = \{B \in \mathcal{B}_X : \mu(T^{-1}B\Delta B) = 0\}$ is the σ -algebra of T -invariant sets (see Glasner (2003)).

We note that in the topological setting we do not have such a decomposition except for distal transformations defined below. For some discussion about this topic see Glasner (1994).

Equicontinuity Versus Measurable Kronecker

Kronecker Systems

For each m.p.s. (X, X, μ, T) we associate a *Koopman operator* $U: L^2(\mu) \rightarrow L^2(\mu)$ such that $Uf(x) = f(Tx)$ for each $f \in L^2(\mu)$. A complex number λ is called an *eigenvalue* of T if there is $0 \neq f \in L^2(\mu)$ such that $Uf = \lambda f$. The function f is called an *eigenfunction* of T corresponding to the eigenvalue λ . It is immediate that T is ergodic if and only if 1 is a simple eigenvalue of T .

An ergodic m.p.s. has *discrete spectrum* if there is an orthonormal basis for $L^2(\mu)$ which consists of eigenfunctions of T . Each ergodic m.p.s. admits a maximal factor with discrete spectrum, called the *Kronecker factor*. Its higher order generalization involving the rotation on nilmanifold will be discussed in the later sections.

If $f \in L^2(\mu)$ is an eigenfunction, then $\text{cl}\{U^n f : n \in \mathbb{Z}\}$ is a compact subset of $L^2(\mu)$. Generally, we say f *almost periodic* if $\text{cl}\{U^n f : n \in \mathbb{Z}\}$ is compact in $L^2(\mu)$. It is well known that the set of all almost periodic functions (denoted by H_c) is spanned by the set of eigenfunctions, and there exists a T -invariant sub- σ -algebra $\mathcal{K}_\mu$ of $\mathcal{B}$ such that $H_c = L^2(X, \mathcal{K}_\mu, \mu)$ (see (Hulse 1982; Zimmer 1976a)). We call $\mathcal{K}_\mu$ the *Kronecker algebra* of (X, X, μ, T) , and $(X, \mathcal{K}_\mu, \mu, T)$ is the *Kronecker factor* of (X, X, μ, T) (Kronecker n.d.). Thus T has discrete spectrum if $H_c = L^2(X, \mathcal{K}_\mu, \mu)$ or equivalently $\mathcal{K}_\mu = \mathcal{X}$.

By the following theorem, an ergodic m.p.s. with discrete spectrum can be viewed as the simplest ergodic system.

Theorem 2.10 (Halmos-von Neumann Theorem). (Halmos and von Neumann 1942) *An ergodic system has discrete spectrum if and only if it is isomorphic to a rotation of a compact monothetic group.*

Equicontinuity

Let (X, T) be a t.d.s. and let $C(X)$ be the vector space of all continuous $\mathbb{C}$ -valued functions on X . Then $(C(X), \|\cdot\|_\infty)$ is a complex Banach space. The Koopman operator $\phi_T: C(X) \rightarrow C(X)$ is defined by $\phi_T(f) = f \circ T$. We say $0 \neq f \in C(X)$ is an eigenfunction for T if there exists $\lambda \in \mathbb{C}$ such that $\phi_T f = \lambda f$. Call λ the eigenvalue for T corresponding to the eigenfunction f .

We say (X, T) has *topological discrete spectrum* if the smallest closed linear subspace of $C(X)$ containing the eigenfunctions of T is $C(X)$, that is, the eigenfunctions span $C(X)$. A transitive t.d.s. (X, T) has topological discrete spectrum if and only if it is minimal and *equicontinuous*, that is, for each $\varepsilon > 0$, there is $\delta > 0$ such that $\rho(x, y) < \delta$ implies that $\rho(T^n x, T^n y) < \varepsilon$ for any $n \in \mathbb{Z}$ (Walters 1982). Note that the property of being equicontinuous does not depend on the choice of the metric d of X . An important fact is that each minimal t.d.s. admits a maximal equicontinuous factor (X_{eq}, T_{eq}) (Ellis 1969; Ellis and Gottschalk 1960). A generalization of this fact to the rotation on nilmanifolds will be discussed in the later section.

By the following theorem, an equicontinuous system can be viewed as the simplest t.d.s.

Theorem 2.11 (Halmos-von Neumann Theorem). (Walters 1982) *Let (X, T) be a t.d.s. Then (X, T) is minimal and equicontinuous if and only if (X, T) is topologically conjugate to a minimal rotation on a compact abelian metric group.*

In the study of topological dynamics, one of the first problems was to find the smallest closed invariant equivalence relation $R(X)$ on (X, T) such that $(X/R(X), T)$ is equicontinuous. A natural candidate for $R(X)$ is the so-called regionally proximal relation $\mathbf{RP}(X)$ introduced by Ellis and Gottschalk (1960). It is a difficult problem to

find conditions under which $\mathbf{RP}(X)$ is an equivalence relation. Starting with Veech (1968), various authors, including MacMahon (1978), Ellis and Keynes (1971), Bronstein (1979), etc., came up with various sufficient conditions for $\mathbf{RP}(X)$ to be an equivalence relation. The generalization of regionally proximal relation to higher orders will be discussed in section “[Structure Theorems and Multiple Ergodic Averages](#).”

Topological Mixing Versus Measurable Mixing

Measurable Mixing

We say an m.p.s. $(X, \mathcal{X}, \mu, T)$ is *strongly mixing* if for any $A, B \in \mathcal{X}$, $\lim_{n \rightarrow \infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B)$.

The stronger notions are the K-system (which will be discussed later) and Bernoulli system which we will not touch, see (Glasner 2003). We remark that a long open question is the Rohlin’s conjecture: if $(X, \mathcal{X}, \mu, T)$ is strongly mixing, is it true that for all $k \in \mathbb{N}$ and $A_0, A_1, \dots, A_k \in \mathcal{X}$

$$\lim_{n_1, \dots, n_k \rightarrow \infty} \mu(A_0 \cap T^{-n_1}A_1 \cap T^{-n_1-n_2}A_2 \cap \dots \cap T^{-n_1-\dots-n_k}A_k) = \mu(A_0) \dots \mu(A_k)?$$

For the progress on the conjecture, see (Kalikow 1984; Ryzhikov 1993; Host 1991; Fayad and Kanigowski 2016; Kanigowski et al. 2017).

Topological Mixing

A t.d.s. (X, T) is (topologically) *strongly mixing* if $N(U, V)$ is cofinite each pair of non-empty open subsets U and V of X , where

$$N(U, V) = \{n \in \mathbb{Z}_+ : U \cap T^{-n}V \neq \emptyset\}.$$

To get a corresponding notion of K-system in topological dynamics, Blanchard (1993) introduced the notion of entropy pair which we will discuss below. We would like to mention that a minimal topological K-system (in the definition of (Huang et al. 2005)) is strongly mixing.

Topological Weak Mixing Versus Measurable Weak Mixing

Topological Weak Mixing

Transitivity and strong mixing are two basic notions of recurrence. There is a notion between them which was introduced by Furstenberg and important to understand the dynamical properties. We say a t.d.s. (X, T) is *weakly mixing* if $(X \times X, T \times T)$ is transitive. Furstenberg (1967) showed that (X, T) is weakly mixing if and only if $\{N(U, V) : U, V \subset X \text{ non-empty open subsets}\}$ is a filter if and only if $N(U, V)$ is thick for each open non-empty subsets U, V of X .

For a minimal system we can say more. Define the upper density of A by

$$\overline{d}(A) = \limsup_{N \rightarrow \infty} \frac{|A \cap \{0, 1, \dots, N-1\}|}{N}.$$

Similarly we define $\underline{d}(A)$. If $\overline{d}(A) = \underline{d}(A) = \alpha$, we then say that the *density* of A is α .

Theorem 2.12 *Let (X, T) be a minimal t.d.s. The following statements are equivalent.*

1. (X, T) is weakly mixing.
2. $N(U, V)$ has density 1 for all $U, V \subset X$ non-empty open subsets (Huang and Ye 2004).
3. The maximal equicontinuous factor is trivial (Auslander 1988).

A nice fact due to Glasner (1994) is that if (X, T) is minimal and weakly mixing, then there is a dense G_δ set X_0 , such that the orbit closure of $(x, \dots, x)$ is dense in X^d under the action of $T \times T^2 \times \dots \times T^d$ for each $x \in X_0$ (for a simple proof see (Moothathu 2010; Kwietniak and Oprocha 2012)). Recently, it was shown by Huang et al. (2019c) that if (X, T) is minimal and weakly mixing, then there is $x \in X$ such that $\{T^{n^2}x : n \in \mathbb{Z}\}$ is dense in X .

Measurable Weak Mixing

Similar to the topological case, there is a notion of weak mixing between ergodicity and strong

mixing. An m.p.s. $(X, \mathcal{X}, \mu, T)$ is *weakly mixing* if $(X \times X, \mathcal{X} \times \mathcal{X}, \mu \times \mu, T \times T)$ is ergodic.

Theorem 2.13 *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic m.p.s. The following statements are equivalent:*

1. $(X, \mathcal{X}, \mu, T)$ is weakly mixing.
2. For every $A, B \in \mathcal{X}$, $\lim_n \mu(A \cap T^{-n}B) = \mu(A)\mu(B)$ outside a set for density zero, that is, for each $\varepsilon < 0$, $\{n \in \mathbb{Z}_+ : |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| < \varepsilon\}$ has density 1.
3. The constants are the only eigenfunctions for T , that is, the Kronecker algebra $\mathcal{K}_\mu$ is trivial.

Topological Mild Mixing Versus Measurable Mild Mixing

Measurable Mild Mixing

We now discuss a recurrence property between weak mixing and strong mixing. It is well known that an m.p.s. $(X, \mathcal{X}, \mu, T)$ is weakly mixing if and only if the product system $(X \times Y, \mathcal{X} \times \mathcal{Y}, \mu \times \nu, T \times S)$ is ergodic for every ergodic m.p.s. $Y = (Y, \mathcal{Y}, \nu, S)$ with $\nu(Y) = 1$. An m.p.s. $(X, \mathcal{X}, \mu, T)$ is called *mildly mixing* if the product system $(X \times Y, \mathcal{X} \times \mathcal{Y}, \mu \times \nu, T \times S)$ is ergodic for every ergodic m.p.s. $Y = (Y, \mathcal{Y}, \nu, S)$ with $\nu(Y) \leq \infty$ (Furstenberg and Weiss 1977). It is easy to see that strong mixing implies mild mixing, and mild mixing implies weak mixing. The following theorem shows that mildly mixing is related to IP^* -sets.

Theorem 2.14 (Furstenberg 1982) *An m.p.s. $(X, \mathcal{X}, \mu, T)$ is mildly mixing if and only if for every $A, B \in \mathcal{X}$ and for each $\varepsilon < 0$, $\{n \in \mathbb{Z}_+ : |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| < \varepsilon\}$ is an IP^* -set.*

We say a measurable function $f \in L^2(X)$ is *rigid* for T if for some sequence $n_k \nearrow \infty$, $T^{n_k}f \rightarrow f$, $k \rightarrow \infty$ in $L^2(X)$. An m.p.s. $(X, \mathcal{X}, \mu, T)$

is *rigid* if for some sequence $n_k \nearrow \infty$, and every $f \in L^2(\mu)$, $T^{n_k}f \rightarrow f$, $k \rightarrow \infty$ in $L^2(\mu)$. An m.d.s. is mildly mixing if and only if it has no nontrivial rigid factor if and only if there are no nonconstant rigid functions in $L^2(X)$ (Furstenberg 1982).

Topological Mild Mixing

For some time, people were looking for a suitable notion of mild mixing in topological dynamics. In the independent works by Huang-Ye and Glasner-Weiss in Huang and Ye (2004) and Glasner and Weiss (2005), the authors defined topological mildly mixing as follows: a t.d.s. (X, T) is *topological mildly mixing* if its product with any transitive system is transitive. The following theorem indicates that topological mildly mixing is also related to IP^* -sets.

Theorem 2.15 *Let (X, T) be a minimal t.d.s. Then it is mildly mixing if and only if $N(U, V)$ is an IP^* -set for each pair of non-empty open subsets U and V of X .*

There are several rigid notions in topological dynamics. A t.d.s. (X, T) is (positively) *rigid* if there is $n_i \rightarrow +\infty$ such that $T^{n_i}x \rightarrow x$, $i \rightarrow \infty$ for each $x \in X$, and it is *uniformly rigid* if $T^{n_i} \rightarrow Id$, $i \rightarrow \infty$ in the uniform topology. We remark that there exists a minimal weakly mixing rigid t.d.s. which is not uniformly rigid (Korner 1987; Glasner and Maon 1989). We also remark that each ergodic measurable rigid system has a topological model which is uniformly rigid (Donoso and Shao 2017).

It was shown (Huang and Ye 2004) that if a minimal system is mildly mixing then it has no nontrivial uniformly rigid factor. It is an open interesting question if the converse is true.

For a given $n \in \mathbb{N}$, we say (X, T) is (positively) *n-rigid* if $(x_1, \dots, x_n)$ is a (positively) recurrent point for $T \times \dots \times T$ (n -times) and each $(x_1, \dots, x_n) \in X^n$. A uniform rigid system has zero topological entropy. It was shown by Weiss (1995) that positively 2-rigidity implies zero entropy and it is an open question if positive or negative 2-rigidity implies zero entropy, see (Weiss 1995; Hochman 2012).

Ellis Semigroup

The enveloping semigroup was introduced by Ellis, and it is a basic tool in the study of dynamical systems.

An *Ellis semigroup* E is a semigroup equipped with a compact Hausdorff topology such that for every $p \in E$ the map $R_p : E \rightarrow E$ defined by $R_p(q) = qp$ is continuous. Let (X, T) be a t.d.s. The *enveloping semigroup* of (X, T) is the closure of $\{T^n : n \in \mathbb{Z}\}$ in X^X which is an Ellis semigroup, denoted by $E(X, T)$. Note that in $E(X, T)$, $p \mapsto Tp$ is continuous, but usually for $p \in E(X, T)$ the map $L_p : E(X, T) \rightarrow E(X, T)$, $q \mapsto pq$ is not continuous. An element v in a semigroup E is an *idempotent* if $v^2 = v$.

Theorem 2.16 (*Ellis-Namakura*). (*Ellis 1969*) *If E is an Ellis semigroup, then there exists an idempotent v in E .*

Let (X, T) be a t.d.s. and $(x, y) \in X^2$. It is a *proximal pair* if

$$\liminf_{n \rightarrow \infty} \rho(T^n x, T^n y) = 0;$$

and it is a *distal pair* if it is not proximal.

Definition 1 A t.d.s (X, T) is *distal* if for all $x \neq y \in X$, (x, y) is a distal pair, that is, $\liminf_{n \rightarrow \infty} \rho(T^n x, T^n y) > 0$.

It is easy to see every equicontinuous system is distal. Let $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$, $(x, y) \mapsto (x + \alpha, x + y)$ for an irrational α . Then this system is minimal distal but not equicontinuous (*Furstenberg 1963*). Using Ellis semigroup, Ellis showed that a t.d.s (X, T) is distal if and only if $E(X, T)$ is a group; and (X, T) is equicontinuous if and only if $E(X, T)$ is a group of homeomorphisms of X (*Ellis 1969*).

For a t.d.s. (X, T) and $x \in X$, the *proximal cell* of x is the set of points y such that (x, y) is proximal. The following fact shows that there always is some minimal point in proximal cell (*Auslander 1988; Ellis 1969*).

Theorem 2.17 [*Auslander-Ellis*] *Let (X, T) be a t.d.s. Then for any $x \in X$ there is a minimal point in the orbit closure of x such that (x, y) is proximal.*

Akin and Kolyada showed that if (X, T) is a weakly mixing t.d.s, then for each $x \in X$, the proximal cell of x is a dense G_δ subset of X (*Akin and Kolyada 2003*). For some related study, see also (*Li et al. 2015b*).

Definition 2 Let (X, T) be a t.d.s. A point $x \in X$ is called a *distal point* if x itself is the only point in its orbit closure which is proximal to x . A minimal t.d.s. (X, T) is called *point distal* if it contains a distal point.

By Auslander-Ellis theorem, a distal point is minimal. If a minimal t.d.s. (X, T) is point distal, then the set of distal point is a dense G_δ subset of X (*Ellis 1973*). If in addition, all points are distal, then (X, T) is distal. If a minimal system is not equicontinuous but it is an almost one-to-one extension of an equicontinuous minimal system (this kind of system is called *almost automorphic*), then it is point distal but not distal. We will introduce the structure theorem of distal and point distal minimal t.d.s. in section “[Structure Theorems and Multiple Ergodic Averages](#).”

Using Theorem 2.17 Furstenberg introduced the notion of central sets. The notion of a central set plays an important role in topological dynamics, and it has very rich combinatorial properties (*Furstenberg 1981*). A subset S of $\mathbb{Z}_+$ is a *central set* if there is a t.d.s. (X, T) , $x \in X$, a minimal point y and a neighborhood U of y such that (x, y) is proximal and $S \supset N(x, U)$. Some important facts: a central set is piecewise syndetic, and contains an IP-set. Note that we say a subset is *piecewise syndetic* if it is the intersection a syndetic subset with a thick subset.

A subset A of $\mathbb{Z}_+$ is called a *dynamical syndetic set*, if there exist a minimal system (Y, S) , $y \in Y$ and an open neighborhood V_y of y such that $A \supset N_S(y, V_y)$. Denote the set of all dynamical syndetic sets by $\mathcal{F}_{ds}$. Let

$$\mathcal{F}_{dps} = \{A \cap B : A \in \mathcal{F}_t, B \in \mathcal{F}_{ds}\}$$

where $\mathcal{F}_t$ is the family of thick sets. It was shown in Huang et al. (2019e) that

Theorem 2.18 $\mathcal{F}_{dps}$ is equal to the family of central sets.

Thus, a central set can be viewed as a dynamical piecewise syndetic set. The following result was shown by Furstenberg (1981).

Theorem 2.19 *Let (X, T) be a t.d.s. and $x \in X$. The following statements are equivalent:*

1. x is distal.
2. (x, y) is recurrent for any recurrent point y of a system (Y, S) .
3. (x, y) is minimal for each minimal point y of a system (Y, S) .
4. $N(x, U)$ is an IP^* set for each neighborhood U of x .

Inspired by the above fact, one can define and study various weak product recurrence properties. For example, one can ask the property of x which satisfies that (x, y) is recurrence for any transitive point y from an M -system (i.e., transitive and with a dense set of minimal points). It was shown that x is then minimal (Dong et al. 2012) and distal (Oprocha and Zhang 2013). Moreover, one can ask the property of x which satisfies that (x, y) is recurrence for any minimal point y . It is easy to see that such an x is not necessarily minimal, and even when x is minimal it is not necessarily distal, see (Glasner and Weiss 2015; Auslander and Furstenberg 1994; Haddad and Ott 2008; Dong et al. 2012; Huang and Ye 2005).

In the measure theoretical setup, Ellis and Nerurkar introduced Enveloping semigroup for an m.p.s. and showed that it plays an important role in ergodic theory (Ellis and Nerurkar 1988, 1989). We refer to Ellis and Nerurkar (1988, 1989) and Glasner (2003) for details.

Disjointness and Weak Disjointness

In number theory one talks about the coprime of two natural numbers. In his seminal paper (Furstenberg 1967) Furstenberg introduced the notion of disjointness both in topological dynamics and ergodic theory in the same spirit. First we discuss disjointness in ergodic theory.

Definition 3 Assume $(X, \mathcal{X}, \mu, T)$ and $(Y, \mathcal{Y}, \nu, S)$ be two m.p.s. We say an invariant measure λ on the product system of X and Y is a *joining* if it projects to μ and ν , respectively. If the only joining is equal to $\mu \times \nu$, we then say that the two m.p.s. are *disjoint*.

Similarly we can define joinings on a collection of m.p.s. $\{(X_i, \mathcal{X}_i, \mu_i, T_i)\}_{i \in I}$. There are certain natural joinings. For example, let

$\phi : (X, \mathcal{X}, \mu, T) \rightarrow (Y, \mathcal{Y}, \nu, S)$ be a homomorphism, then the so-called *graph joining* $gr(\mu, \phi)$ is defined by $gr(\mu, \phi)(A \times B) = \mu(A \cap \phi^{-1}B)$. One may find more examples of joinings in Glasner (2003, Chap. 6).

Furstenberg showed (Furstenberg 1967) that if two m.p.s. are disjoint then one of them has zero entropy and he was able to characterize the systems as follows:

Theorem 2.20 *An m.p.s. is measurable K-system if and only if it is disjoint from every zero entropy m.p.s. An ergodic system is weakly mixing if and only if it is disjoint from every ergodic discrete spectrum system.*

One may find definitions of K-system and u.p.e. in the next section. From this theorem one sees that two disjoint systems should have different properties. Nowadays the notion of joining is a very useful tool in the study of ergodic theory.

It is easy to see that two disjoint systems do not have common nontrivial factors. It is a question for a while if the converse is true. The first counterexample was provided by Rudolph in Rudolph (1979). The negative answer to Furstenberg's question and the consequent works on joinings and disjointness show that in order to study the relationship between two dynamical systems it is necessary to know all the possible joinings of the two systems and to understand the nature of these joinings. We refer to Glasner's book (Glasner 2003) for the discussion.

Now we turn to the topological case.

Definition 4 If (X, T) and (Y, S) are two t.d.s., we say $J \subset X \times Y$ is a *joining* of X and Y if J is a non-empty closed $T \times S$ -invariant set and is projected onto X and Y . If the only joining is equal to $X \times Y$, we then say that (X, T) and (Y, S) are *disjoint* or $(X, T) \perp (Y, S)$ or $X \perp Y$.

The examples of two minimal t.d.s. which have no common nontrivial factor and are not disjoint were found by Glasner and Weiss (1983) and Lindenstrauss (1995).

Compared with Theorem 2.20, we have that a minimal u.p.e. system is disjoint from minimal zero entropy systems (Blanchard 1993) and a

topologically weakly mixing system is disjoint from minimal equicontinuous systems.

Let $\mathcal{M}$ be the collection of all minimal systems. Furstenberg (1967) showed that if two systems are disjoint then one of them is minimal (the other one must have dense recurrent points (Huang and Ye 2005)). Thus the most natural question is

Question E in Furstenberg (1967): Characterize $\mathcal{M}^\perp$.

Through years of efforts of many researchers this question was answered finally in Huang et al. (2019e). Before we state the result we first describe some sufficient and necessary conditions. We remark that Furstenberg showed (Furstenberg 1967) that a weakly mixing t.d.s. with dense periodic points is in $\mathcal{M}^\perp$.

Recall that a t.d.s. is an M -system if it is transitive and the set of minimal points is dense. It is easy to show that a t.d.s. (X, T) is an M -system if and only if for any transitive point x and its neighborhood U , $N(x, U)$ is piecewise syndetic. First we state the following result.

Theorem 2.21 *Let (X, T) be a transitive t.d.s. If $(X, T) \perp \mathcal{M}$, then (X, T) is a weakly mixing M -system (Huang and Ye 2005). Each weakly mixing system with a dense set of distal points is in $\mathcal{M}^\perp$ (Oprocha 2010; Dong et al. 2012).*

Now we present an answer to the question in the transitive case, for the general case see (Huang et al. 2019e). We remark that the sufficient condition was proved in Oprocha (2019).

Theorem 2.22 *Let (X, T) be a transitive t.d.s. Then $(X, T) \perp \mathcal{M}$ if and only if (X, T) is weakly mixing and there is a countable dense subset D of X such that for each non-empty open subset U of X and each central set $S = A \cap B$, we can find $x = x(B) \in D \cap U$ with $N_T(x, U) \cap S \neq \emptyset$, where A is thick and B is dynamical syndetic.*

Now we turn to the notion of weak disjointness. We say two systems are *weakly disjoint* if the product system is transitive in the topological settings or ergodic in the measure theoretical setup. Weak disjointness is very useful to define new dynamical properties. We have defined mild mixing in the previous sections in such a way.

In the measure-theoretical setup, it is easy to check that an m.p.s. is weakly mixing if and only if it is weakly disjoint from every ergodic m.p.s. if and only if it is weakly disjoint from every weakly mixing m.p.s.

In the topological situation, we say a t.d.s. is *scattering* if it is weakly disjoint from every minimal t.d.s. Originally the notion of scattering was defined by using the complexity of an open cover, the equivalence to the above definition was given in Blanchard et al. (2000). Using weak disjointness one can define many other scattering properties, see (Huang and Ye 2002b, 2004).

For more details about disjointness, we refer to Glasner (2003), Rudolph (1990), and de la Rue (2012).

Chaos and Complexity: Topological Versus Measurable

Chaos and complexity theory in both of topological dynamics and ergodic play an important role in the study. Below we only pick some of them to present the similarity in both theories.

First we consider the notion of sensitivity. We say a t.d.s. (X, T) is *sensitive* if there is $\delta > 0$ such that for each $x \in X$ and $\varepsilon > 0$ there are y with $\rho(x, y) < \varepsilon$ and $n \in \mathbb{N}$ with $\rho(T^n x, T^n y) > \delta$. We have

Theorem 2.23 (Akin et al. 1993; Auslander and Yorke 1980) *A minimal t.d.s. is either sensitive or equicontinuous.*

Let (X, T) be a t.d.s., $\mathcal{B}_X$ be the Borel σ -algebra and μ be an invariant measure for T . Then T is said to be *sensitive for μ* , if there is $\delta > 0$ such that for any $A \in \mathcal{B}_X$ with $\mu(A) > 0$ we can find $x, y \in A$ and $n \geq 0$ with $\rho(T^n x, T^n y) > \delta$. For a similar notion, see (Cadre and Jacob 2005; James et al. 2008). We have:

Theorem 2.24 (Huang et al. 2011a) *Let (X, T) be a t.d.s. and p be an ergodic measure on (X, T) . Then the ergodic m.p.s. $(X, \mathcal{B}_X, \mu, T)$ is either sensitive for μ or has discrete spectrum.*

Sensitivity is only one aspect of chaotic behaviors. Nowadays, there are many definitions of chaos, for example, Li-Yorke's chaos (Li and Yorke 1975), Devaney's chaos (Devaney 1989),

positive entropy, distributional chaos (Schweizer and Smítal 1994). For the papers dealing with the relationship among the definitions, see (Blanchard et al. 2002a; Huang et al. 2014; Downarowicz 2014; Huang and Ye 2002a; Li and Ye 2016).

Now we consider complexity. There are many ways to define complexity. Below we only choose some of them which relates some new distances induced from the original one. Let (X, T) be a t.d.s. For $n \in \mathbb{N}$ and $x, y \in X$, define the Bowen's metric

$$\rho_n(x, y) = \max \{ \rho(T^i x, T^i y) : i = 0, 1, \dots, n-1 \}.$$

We say that $W(\varepsilon, n)$ - spans B if for any $x \in B$ there is $y \in W$ such that $\rho_n(x, y) < \varepsilon$, where $\varepsilon > 0$ and $n \in \mathbb{N}$. A subset of X is said to be a (T, ε, n) -span if it (ε, n) -spans X . Let $\text{span}(\varepsilon, n)$ denote the smallest cardinalities of all (T, ε, n) -spans. We say that X has *bounded complexity with respect to* $\{\rho_n\}$ if for every $\varepsilon > 0$, $\{\text{span}(\varepsilon, n)\}_n$ is a bounded sequence.

Theorem 2.25 *Let (X, T) be a t.d.s. Then X has bounded complexity with respect to $\{\rho_n\}$ if and only if it is equicontinuous.*

Let (X, T) be a t.d.s. and μ be an invariant measure on (X, T) . For $n \in \mathbb{N}$ and $\varepsilon > 0$, let $\text{span}_\mu(n, \varepsilon)$ be

$$\min \left\{ \#(F) : F \subset X \text{ and } \mu \left(\bigcup_{x \in F} B_{\rho_n}(x, \varepsilon) \right) > 1 - \varepsilon \right\}.$$

Recall that this is the same notion defined in Katok (1980) by Katok. We say that μ has *bounded complexity with respect to* $\{\rho_n\}$ if for every $\varepsilon > 0$, $\{\text{span}_\mu(n, \varepsilon)\}_n$ is a bounded sequence. We say T is μ -equicontinuous if for each $\varepsilon > 0$ there is a compact subset K with $\mu(K) > 1 - \varepsilon$ and $T|_K$ is equicontinuous. We have:

Theorem 2.26 (Huang et al. 2019a) *Let (X, T) be a t.d.s. and μ be an invariant measure on (X, T) . Then μ has bounded complexity with respect to $\{\rho_n\}$ if and only if T is μ -equicontinuous.*

Let (X, T) be a t.d.s. For $n \in \mathbb{N}$ and $x, y \in X$, we can define the so-called *mean metric*

$$\bar{\rho}_n(x, y) = \frac{1}{n} \sum_{i=0}^{n-1} \rho(T^i x, T^i y)$$

We may define the complexity with respect to the mean metric. The bounded complexity is related to mean equicontinuity and μ -mean equicontinuity. For related papers, see (Huang et al. 2019a; Li et al. 2015a; García-Ramos 2017; Qiu and Zhao 2020; Yu 2019). We remark that the complexity for an invariant measure can be used to the study of Sarnak's conjecture (Huang et al. 2019f) (see section “Sarnak Conjecture”).

Meet Together

Let (X, T) be a t.d.s. and $\mathcal{B}_X$ be the Borel σ -algebra. Denote by $\mathcal{M}(X)$ the set of all Borel probability measures on $\mathcal{M}(X, T)$ the set of T -invariant measures, and $\mathcal{M}^e(X, T)$ the set of ergodic measures. Note that the ergodic measures are characterized as the extreme points of the Choquet simplex $\mathcal{M}(X, T)$. Then both $\mathcal{M}(X)$ and $\mathcal{M}(X, T)$ are convex, compact metric spaces endowed with the weak*-topology. A useful fact is that for any $x \in X$, any weak*-limit point of $\frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i x}$ is an invariant measure for T . A Borel probability measure $\mu \in \mathcal{M}(X, T)$ induces an m.p.s. $(X, \mathcal{B}_X, \mu, T)$. A t.d.s. (X, T) is *uniquely ergodic* if there is only one invariant measure for T , that is, $\mathcal{M}(X, T)$ consists of a single element. A uniquely ergodic minimal t.d.s. (X, T) is also called *strictly ergodic*.

Let $(X, \mathcal{X}, \mu, T)$ be an ergodic m.p.s. We say that $(\hat{X}, \hat{T})$ is a *topological model* (or just a *model*) for $(X, \mathcal{X}, \mu, T)$ if $(\hat{X}, \hat{T})$ is a t.d.s. and there exists an invariant probability measure $\hat{\mu}$ on the Borel σ -algebra $\mathcal{B}_{\hat{X}}$ such that the systems $(X, \mathcal{X}, \mu, T)$ and $(\hat{X}, \mathcal{B}_{\hat{X}}, \hat{\mu}, \hat{T})$ are measured theoretically isomorphic. It is well known that each m.p.s. has a topological model (Furstenberg 1981). Weiss (1989) showed the following nice result: there exists a minimal t.d.s. (X, T) with the property that for every aperiodic ergodic m.p.s. $(Y, \mathcal{Y}, \nu, S)$ there exists $\mu \in \mathcal{M}(X, T)$ such that $(Y, \mathcal{Y}, \nu, S)$ and $(X, \mathcal{B}_X, \mu, T)$ are isomorphic. At the same time, there is no universal topological model in the class of entropy zero systems (Serafin 2013).

The pioneering work on topological model was done by Jewett (1969/1970) and Krieger (1970/1971). That is

Theorem 2.27 (*Jewett-Krieger*). *Let $(X, \mathcal{X}, T, \mu)$ be an ergodic m.p.s. Then there is a minimal uniquely ergodic t.d.s. $(Y, \mathcal{Y})$ with invariant measure ν such that $(X, \mathcal{X}, \mu, T)$ is isomorphic to $(Y, \mathcal{B}_Y, \nu, S)$.*

We note that one can add some additional properties to the topological model. For example, in Lehrer (1987) Lehrer showed that the strictly ergodic model can be required to be a topological (strongly) mixing system in addition. An ergodic system has a doubly minimal (i.e., the orbit of any point $(x, y) \in X^2$ is dense provided they are not in the same orbit) model if and only if it has zero entropy (Weiss 1995); and an ergodic system has a strictly ergodic, UPE (uniform positive entropy) model if and only if it has positive entropy (Glasner and Weiss 1994). Note that not any dynamical properties can be added in the uniquely ergodic models. For example, Lindenstrauss showed that every ergodic measurable distal system $(X, \mathcal{X}, \mu, T)$ has a minimal topologically distal model (Lindenstrauss 1999). This topological model need not, in general, be uniquely ergodic. Weiss (1985) generalized the theorem of Jewett-Krieger to the relative case which will be discussed below. For more information we refer to Glasner and Weiss (2005).

Entropy Theory

Entropy is one of the measurements of the complexity of dynamical systems. It is an important conjugacy or isomorphic invariant. The concept of entropy was invented by Clausius in 1854, and Shannon introduced it into information theory in 1948. In 1958 Kolmogorov carried it over to ergodic theory. One can find a brief history of entropy in the recent book (Downarowicz 2011). There are lots of excellent books introducing the classical theory of entropy such as (Downarowicz 2011; Glasner 2003; Rudolph 1990; Walters 1982). In this entry we mainly focus on the so-called local entropy theory, starting from the pioneer work of Blanchard in the early 1990s. In such a theory one studies the local properties of

tuples in the product space and it can be used to investigate the global properties of the systems. For example, by using the properties of entropy pairs, one obtains the existence of the maximal zero entropy factor of any t.d.s. (topological Pinsker factor). For related survey, see (Glasner and Ye 2009; Oprocha and Zhang 2014).

Entropy: Topological Versus Measurable

In 1958 Kolmogorov firstly introduced the entropy of an m.p.s. by generating partition (Kolmogorov 1958, 1959). Sinai subsequently found a natural way to make this notion better-behaved by observing that generators maximize entropy relative to a partition among all partitions with finite entropy, and formulated the definition that has become the standard one (Sinai 1959). Later the topological entropy was introduced by Adler et al. who formulated it in terms of open covers (Adler et al. 1965). Equivalent definition based on separated and spanning set was independently given by Bowen (1971) and Dinaburg (1971).

Now we define entropy. Let (X, T) be a t.d.s. A *cover* of X is a family of subsets of X , whose union is X . A *partition* of X is a cover of X whose elements are pairwise disjoint. Given two covers $\mathcal{U}, \mathcal{V}$ of X , $\mathcal{U}$ is said to be finer than $\mathcal{V}$ (denoted by $\mathcal{U} \succcurlyeq \mathcal{V}$ or $\mathcal{V} \preccurlyeq \mathcal{U}$) if each element of $\mathcal{U}$ is contained in some element of $\mathcal{V}$; set $\mathcal{U} \vee \mathcal{V} = \{U \cap V : U \in \mathcal{U}, V \in \mathcal{V}\}$ and $T^{-i}\mathcal{U} = \{T^{-i}U : U \in \mathcal{U}\}$ for $i \in \mathbb{Z}_+$. Denote by $N(\mathcal{U})$ the minimal cardinality among all cardinalities of subcovers of $\mathcal{U}$.

Definition 5 Let (X, T) be a t.d.s. and $\mathcal{U}$ be a finite open cover of X . The *topological entropy* of $\mathcal{U}$ is defined by

$$h_{\text{top}}(T, \mathcal{U}) = \lim_{n \rightarrow +\infty} \frac{1}{n} \log N(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}).$$

The *topological entropy* of (X, T) is $h_{\text{top}}(T) = \sup_{\mathcal{U}} h_{\text{top}}(T, \mathcal{U})$, where supremum is taken over all finite open covers of X .

Note that $\{\log N(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U})\}_{n=1}^{\infty}$ is a sub-additive sequence and hence $h_{\text{top}}(T, \mathcal{U})$ is well defined.

Let $(X, \mathcal{X}, \mu, T)$ be an m.p.s. and $\mathcal{P}_X$ be the set of finite measurable partitions of X . Suppose $\xi \in \mathcal{P}_X$ and an σ -algebra $\mathcal{A} \subset \mathcal{X}$. The *entropy of ξ* (with respect to $\mathcal{A}$), written $H_\mu(\xi)$ (resp. $H_\mu(\xi|\mathcal{A})$), is defined by the formula

$$H_\mu(\xi) = - \sum_{A \in \xi} \mu(A) \log \mu(A)$$

and $H_\mu(\xi|\mathcal{A}) = \sum_{A \in \xi} \int_X -\mathbb{E}_\mu(1_A|\mathcal{A}) \log \mathbb{E}_\mu(1_A|\mathcal{A}) d\mu$, where $\mathbb{E}_\mu(1_A|\mathcal{A})$ is the expectation of 1_A with respect to $\mathcal{A}$.

Definition 6 Let $(X, \mathcal{X}, \mu, T)$ be an m.p.s. and $\xi \in \mathcal{P}_X$. The *entropy of ξ* is defined by

$$h_\mu(T, \xi) = \limsup_{n \rightarrow +\infty} \frac{1}{n} H_\mu(\bigvee_{i=0}^{n-1} T^{-i} \xi).$$

And the *entropy of $(X, \mathcal{X}, T, \mu)$* is

$$h_\mu(T) = \sup_{\alpha \in \mathcal{P}_X} h_\mu(T, \alpha).$$

It is well known that $h_{\text{top}}(T)$ and $h_\mu(T)$ are conjugacy and isomorphism invariants respectively, and they have been the most successful invariant so far in t.d.s. and m.p.s. In 1958 Kolmogorov asked if entropy is a complete isomorphic invariant on the collection of Bernoulli shifts, see (Glasner 2003). This was answered affirmatively by Ornstein in 1970 (Ornstein 1970).

The basic relationship between topological entropy and measure-theoretic entropy is given by the variational principle (Goodwyn 1969; Goodman 1971; Misiurewicz 1976).

Theorem 3.1 (The variational principle). *Let (X, T) be a t.d.s. Then*

$$\begin{aligned} h_{\text{top}}(T) &= \sup \{ h_\mu(T) : \mu \in \mathcal{M}(X, T) \} \\ &= \sup \{ h_\mu(T) : \mu \in \mathcal{M}^e(X, T) \}. \end{aligned}$$

Measurable and Topological K -Systems

In this subsection we discuss an important class of m.p.s. in ergodic theory: K -systems, and its topological analogy. Introduced by Kolmogorov (1958), it is an isomorphism invariant version of

earlier regularity notions for random processes: present becomes asymptotically independent of all sufficiently long past.

Definition 7 An m.p.s. $(X, \mathcal{X}, \mu, T)$ is called a *Kolmogorov system*, or a *K-system*, if there is a sub σ -algebra $\mathcal{K} \subset \mathcal{X}$ with the following properties:

1. $\mathcal{K} \subset T\mathcal{K}$,
2. $\bigvee_{n=0}^{\infty} T^n \mathcal{K} = \mathcal{X} \pmod{\mu}$,
3. $\bigcap_{n=0}^{\infty} T^{-n} \mathcal{K} = \{X, \emptyset\} \pmod{\mu}$.

A K -system completely differs from zero entropy systems, see Theorem 2.20. It can be viewed as the most complicated systems in the language of entropy. A fundamental result is that an m.p.s. is K -system if and only if it has *completely positive entropy* (each nontrivial factor has positive entropy) if and only if every measurable partition by two nontrivial elements has positive entropy if and only if every measurable partition by finite nontrivial elements has positive entropy (Pinsker 1960; Rohlin 1967).

Now we consider t.d.s. Using the first two conditions above Blanchard (1992) introduced the notion of c.p.e. and u.p.e. in t.d.s. as an analogue of the K -system in ergodic theory.

Definition 8 A t.d.s. (X, T) has *completely positive entropy* (c.p.e. for short) if each nontrivial factor has positive entropy and *uniform positive entropy* (u.p.e. for short) if every cover by two nontrivial open sets has positive entropy.

Blanchard (1992) showed that u.p.e. implies weak mixing and c.p.e. implies the existence of an invariant measure with full support. The topic on the relative notion of c.p.e. and u.p.e. can be found in Glasner and Weiss (1995b) and Huang et al. (2007).

Definition 9 (Huang and Ye 2006) Let (X, T) be a t.d.s. (X, T) has *uniform positive entropy of order n* (u.p.e. of order n , for short), if any cover of X by n non-dense open sets has positive topological entropy. We say (X, T) has u.p.e. of all orders or it is *topological K* if it has u.p.e. of order n for every $n \geq 2$.

Clearly, u.p.e. of order 2 is just u.p.e. It is shown that a u.p.e. system is mildly mixing (Huang and Ye 2006); and any minimal topological K is strongly mixing (Huang et al. 2005).

Huang and Ye (2006) answered several open questions concerning the nature of u.p.e. and c.p.e. Namely, they showed that u.p.e. of order n does not imply u.p.e. of order $n + 1$ for each $n \geq 2$ (answering a question by Host (Glasner and Weiss 1995b)); there is a transitive diagonal system which does not have u.p.e. (of order 2) (Blanchard 1993, Question 1); there is a u.p.e. (of order 2) system having no ergodic measure with full support (Blanchard 1992, Question 2).

In Glasner and Weiss (1994) Glasner and Weiss showed that if a t.d.s. admits a K-measure with full support, then it has u.p.e; and there is a minimal u.p.e. system that is universal for any ergodic m.p.s. with positive entropy. Above results were extended to u.p.e. of all orders in Huang and Ye (2006). Particularly, the authors proved that if a t.d.s. admits an invariant K-measure with full support, then it has u.p.e. of all orders, and a t.d.s. (X, T) has u.p.e. of all orders if and only if there is an invariant measure $\mu \in \mathcal{M}(X, T)$ such that for each partition α of X by finite non-dense Borel sets one has $h_\mu(T, \alpha) > 0$.

Measurable and Topological Entropy Tuples

Now we consider entropy tuples which began with Blanchard's work for entropy pairs (Blanchard 1993).

Entropy N-Tuples

Given a t.d.s. (X, T) and an integer $n \geq 2$, the n -th product system is the t.d.s. $(X^n, T^{(n)})$, where $T^{(n)} = T \times \dots \times T$ (n times). Let $\mathcal{B}_X$ be the Borel σ -algebra and $\mathcal{B}_X^{(n)}$ be the Borel σ -algebra of X^n . The diagonal of X^n is denoted by $\Delta_n(X) = \{(x_i)_1^n \in X^n : x_1 = \dots = x_n\}$. Let $(x_i)_1^n \in X^n$. A finite cover of X , $\mathcal{U}$, is said to be an *admissible cover* with respect to $(x_i)_1^n$, if all the points x_i do not belong to the closure of the same element U of $\mathcal{U}$. Analogously we define admissible partitions with respect to $(x_i)_1^n$.

Definition 10 (Glasner and Weiss 1994; Huang and Ye 2006). Let (X, T) be a t.d.s. An n -tuple $(x_i)_1^n \in X^n$, $n \geq 2$, is called an *entropy n -tuple* if for some $1 \leq i \neq j \leq n$, $x_i \neq x_j$, and for any admissible open cover $\mathcal{U}$ with respect to $(x_i)_1^n$, $h_{top}(T, \mathcal{U}) > 0$. Entropy 2-tuples are called *entropy pairs*.

We denote by $E_n(X, T)$ the set of entropy n -tuples. Then following the ideas of Blanchard (1993) we have

1. If $\mathcal{U} = \{U_1, \dots, U_n\}$ is an open cover of X with $h_{top}(T, \mathcal{U}) > 0$, then for all $1 \leq i \leq n$ there exists $x_i \in U_i^c$ such that $(x_i)_1^n$ is an entropy n -tuple.
2. $E_n(X, T) \cup \Delta_n(X)$ is a closed $T^{(n)}$ -invariant subset of X^n .
3. Let $\pi : (X, T) \rightarrow (Y, S)$ be a factor map. Then $\pi^{(n)}(E_n(X, T) \cup \Delta_n(X)) = E_n(Y, S) \cup \Delta_n(Y)$.

It follows from (1) that (X, T) has positive entropy if and only if $E_2(X, T) \neq \emptyset$; (X, T) is u.p.e. of order n if and only if for every point $(x_i)_1^n \in X^n$ not on the diagonal $\Delta_n(X)$ is entropy n -tuple.

Definition 11 Let $(X, \mathcal{X}, \mu, T)$ be an m.p.s. The *Pinsker σ -algebra* is $\mathcal{P}_\mu = \{A \in \mathcal{X} : h_\mu(T, \{A, X \setminus A\}) = 0\}$. The corresponding factor is called the *Pinsker factor* of $(X, \mathcal{X}, \mu, T)$.

The Pinsker factor of $(X, \mathcal{X}, \mu, T)$ is the largest zero entropy factor. By Rohlin-Sinai's theorem, an m.p.s. $(X, \mathcal{X}, \mu, T)$ is K-system if and only if the Pinsker factor is trivial, that is, $\mathcal{P}_\mu = \{X, \emptyset\} \pmod{\mu}$ (Rohlin and Sinai 1961). Now we define the topological Pinsker factor.

Definition 12 Let (X, T) be a t.d.s. The *topological Pinsker factor* is the maximal factor with zero topological entropy.

Blanchard and Lacroix (1993) showed that for a t.d.s. (X, T) , the smallest closed invariant equivalence relation containing $E_2(X, T)$ induces the topological Pinsker factor. For the relative version, see (Park and Siemaszko 2001).

In Blanchard et al. (1995) the authors introduced the notion of entropy pairs for a measure and their notion cannot be directly generalized for n -tuples when $n > 2$. Here we will give a definition of entropy n -tuples for a measure, which was introduced by Huang and Ye in Huang and Ye (2006) and is the same as the notion of entropy pairs for a measure when $n = 2$.

Definition 13 Let (X, T) be a t.d.s. and $\mu \in \mathcal{M}(X, T)$. An n -tuple $(x_i)_1^n \in X^{(n)}$, $n \geq 2$, is called an entropy n -tuple for μ if for some for some $1 \leq i \neq j \leq n$, $x_i \neq x_j$, and for any admissible Borel partition α with respect to $(x_i)_1^n$, $h_\mu(T, \alpha) > 0$.

Let $\mu \in \mathcal{M}(X, T)$ and $\mathcal{P}_\mu$ be the Pinsker σ -algebra of $(X, \mathcal{B}_X, \mu, T)$. Define the conditional independent joining $\lambda_n(\mu)$ on $(X^{(n)}, \mathcal{B}_X^n, T^{(n)})$ by

$$\lambda_n(\mu) \left(\prod_{i=1}^n A_i \right) = \int_X \prod_{i=1}^n \mathbb{E}(1_{A_i} | \mathcal{P}_\mu) d\mu,$$

where $A_i \in \mathcal{B}_X$, $i = 1, \dots, n$. The following result shows that the set of entropy n -tuples for an invariant measure is in the support of $\lambda_n(\mu)$ and we remark that the case $n = 2$ was proved in Glasner (1997) and the general case was proved in Huang and Ye (2006).

Theorem 3.2 Let (X, T) be a t.d.s., $\mu \in \mathcal{M}(X, T)$ and $n \geq 2$. Then

$$E_n^\mu(X, T) = \text{supp}(\lambda_n(\mu)) \setminus \Delta_n(X).$$

By Theorem 3.2 it is clear that $h_\mu(T) = 0$ if and only if $E_2^\mu(X, T) = \emptyset$; $E_n^\mu(X, T) \cup \Delta_n(X)$ is a closed and $T^{(n)}$ -invariant subset of X^n .

Local Variational Principles – The Connection of Two Kinds of Tuples

To study the relationship between the two kinds of entropy pairs or tuples, one needs a local version of the variational principle. Blanchard, Glasner, and Host obtained the following:

Theorem 3.3 (Blanchard et al. 1997) For a given t.d.s. (X, T) and a finite open cover $\mathcal{U}$ of X there is an invariant measure $\mu \in \mathcal{M}(X, T)$ with $\inf_\alpha h_\mu(T, \alpha) \geq h_{\text{top}}(T, \mathcal{U})$, where the infimum is taken over all finite Borel partitions of X which are finer than $\mathcal{U}$.

In Huang and Ye (2006) Huang and Ye showed that, if $\mu \in \mathcal{M}(X, T)$ and $h_\mu(T, \alpha) > 0$ for each finite Borel partition α of X which is finer than $\mathcal{U}$, then $\inf_\alpha h_\mu(T, \alpha) > 0$ and $h_{\text{top}}(T, \mathcal{U}) > 0$ providing some kind of converse statement of Theorem 3.3.

To study the question whether $\inf_\alpha h_\mu(T, \alpha) = h_{\text{top}}(T, \mathcal{U})$ for a given finite open cover $\mathcal{U}$, Romagnoli (2003) introduced the following entropy for covers

$$h_\mu^+(T, \mathcal{U}) = \inf_{\alpha \succ \mathcal{U}} h_\mu(T, \alpha) \text{ and}$$

$$h_\mu(T, \mathcal{U}) = \lim_{n \rightarrow +\infty} \frac{1}{n} \inf_{\alpha \succ \bigvee_{i=0}^{n-1} T^{-i} \mathcal{U}} H_\mu(\alpha),$$

where α is a finite Borel partition of X . In fact they are the same: $h_\mu^+(T, \mathcal{U}) = h_\mu(T, \mathcal{U})$ (Huang et al. 2006).

Theorem 3.4 (The local variational principle). (Romagnoli 2003; Glasner and Weiss 2005) Let (X, T) be a t.d.s. and $\mathcal{U}$ be a finite open cover of X . Then

$$\begin{aligned} \max_{\mu \in \mathcal{M}(X, T)} h_\mu^+(T, \mathcal{U}) &= \max_{\mu \in \mathcal{M}(X, T)} h_\mu(T, \mathcal{U}) \\ &= h_{\text{top}}(T, \mathcal{U}). \end{aligned}$$

Note that the classical variational principle follows from the local ones by note that $h_\mu(T) = \sup_{\mathcal{U}} h_\mu(T, \mathcal{U})$, where supremum is taken over all finite open covers of X .

Topological entropy tuples are closely related to measurable ones. For each invariant measure the set of entropy pairs for a measure is contained in the set of entropy pairs (Blanchard et al. 1995); the converse is also valid (Blanchard et al. 1997). A variational relation of entropy n -tuple is as follows:

Theorem 3.5 (Huang and Ye 2006) *Let (X, T) be a t.d.s. If $\mu \in \mathcal{M}(X, T)$, then $E_n(X, T) \supseteq E_n^\mu(X, T)$ for each $n \geq 2$ and there exists $\mu \in \mathcal{M}(X, T)$ such that $E_n(X, T) = E_n^\mu(X, T)$ for each $n \geq 2$.*

We remark that Blanchard et al. (1997) constructed a t.d.s. and an entropy pair for that system, which is not a metric entropy pair for any ergodic measure. Also the property that the product of u.p.e. of order n (resp. of all orders) systems is again u.p.e. of order n (resp. of all orders) was proved by Glasner (1997) for $n = 2$ and by Huang and Ye (2006) for the general case.

The Ergodic Decomposition

Let (X, T) be a t.d.s. Let $\mu \in \mathcal{M}(X, T)$ and $\mu = \int_X \mu_x d\mu(x)$ be its ergodic decomposition. The ergodic decomposition of μ also gives an ergodic decomposition of the μ -entropy of a $\alpha \in \mathcal{P}_X$: $h_\mu(T, \alpha) = \int_X h_{\mu_x}(T, \alpha) d\mu(x)$ (Denker et al. 1976). This property also holds for $h_\mu(T, \mathcal{U})$ for any finite Borel cover $\mathcal{U}$ of X .

Theorem 3.6 (Huang and Ye 2006) *Let (X, T) be a t.d.s., $\mu \in \mathcal{M}(X, T)$ and $\mathcal{U}$ be a finite Borel cover of X . If $\mu = \int_X \mu_x d\mu(x)$ is the ergodic decomposition of μ then*

$$h_\mu(T, \mathcal{U}) = \int_X h_{\mu_x}(T, \mathcal{U}) d\mu(x)$$

Given a finite open cover $\mathcal{U}$ of a t.d.s. (X, T) , we know that there exists $\nu \in \mathcal{M}^e(X, T)$ with $h_\nu(T, \mathcal{U}) = h_{\text{top}}(T, \mathcal{U})$ by Theorem 3.6 and the local variational principle. Moreover, it was showed that the entropy map $\mu \in \mathcal{M}(X, T) \mapsto h_\mu(T, \mathcal{U})$ is upper semicontinuous (Huang et al. 2011b).

The following result discloses the relation of entropy tuples for an invariant measure and entropy tuples for ergodic measures in its ergodic decomposition, which was proven for $n = 2$ in Blanchard et al. (1997) and in general (Huang and Ye 2006).

Theorem 3.7 *Let (X, T) be a t.d.s. and $\mu \in \mathcal{M}(X, T)$ with $\mu = \int_X \mu_x d\mu(x)$ the ergodic decomposition of μ . Then*

1. *for μ -a.e. $x \in X$, $E_n^{\mu_x}(X, T) \subseteq E_n^\mu(X, T)$ for each $n \geq 2$.*
2. *$(x_i)_1^n \in E_n^\mu(X, T)$, then for every neighborhood V of $(x_i)_1^n$,*

$$\mu(\{x \in X : V \cap E_n^{\mu_x}(X, T) \neq \emptyset\}) > 0.$$

Thus we can choose $X_0 \in \mathcal{B}_X$ such that $\mu(X_0) = 1$ and $\cup \{E_n^{\mu_x}(X, T) : x \in X_0\} \setminus \Delta_n(X) = E_n^\mu(X, T)$

Weak Horseshoe

In Huang and Ye (2006) Huang and Ye obtained the following equivalent characterization of topological entropy n -tuples. Let (X, T) be a t.d.s. and $n \geq 2$. Then $(x_1, \dots, x_n) \in E_n(X, T)$ if and only if for any neighborhood $U_1 \times \dots \times U_n$ of $(x_1, \dots, x_n)$, there exists a positive density subset $S = \{s_1 < s_2 < \dots\}$ of $\mathbb{Z}_+$ such that $\cap_{i=1}^\infty T^{-s_i} U_{t(i)} \neq \emptyset$ for any $t \in \{1, 2, \dots, n\}^S$ (see also (Kerr and Li 2007, 2009)).

Motivated by this, let J be a subset of $\mathbb{Z}_+$ we say that (X, T) has a *weak horseshoe with an interpolating set J* if there exist two disjoint closed subsets U_0, U_1 of X such that for any $t \in \{0, 1\}^J$, $\cap_{j \in J} T^{-j} U_{t(j)} \neq \emptyset$, that is, there exists $x_t \in X$ such that $T^j(x_t) \in U_{t(j)}$ for any $j \in J$ (Huang and Lu 2017). If the subset J has positive density, then we say that (X, T) has a *weak horseshoe*.

Theorem 3.8 (Huang and Ye 2006) *A t.d.s. has positive topological entropy if and only if it has a weak horseshoe.*

This result has many applications. Note that the result was proved by Glasner and Weiss in Glasner and Weiss (1995a) for symbolic dynamics, and was extended to countable amenable group actions by Kerr and Li in Kerr and Li (2007). Recently, in Huang and Lu (2017) Huang and Lu studied the complicated dynamics of infinite dimensional random dynamical systems and showed that in this setting positive topological entropy also implies the existence of weak horseshoes.

Sequence Entropy: Topological Versus Measurable

The sequence entropy, of an m.p.s. for a sequence of $\mathbb{Z}_+$ was introduced by Kushnirenko in 1967 (Kushnirenko 1967). Later, the topological sequence entropy was investigated by Goodman in 1974 (Goodman 1974).

Now we define sequence entropy. Denote by $\mathcal{F}_{inf}$ the set of all increasing sequences of $\mathbb{Z}_+$. Let $S = \{0 \leq t_1 \leq t_2 \leq \dots\} \in \mathcal{F}_{inf}$ and $\mathcal{U}$ be a finite open cover of X . The *topological sequence entropy of $\mathcal{U}$ with respect to (X, T) along S* is defined by

$$h_{top}^S(T, \mathcal{U}) = \limsup_{n \rightarrow +\infty} \frac{1}{n} \log N(\bigvee_{i=1}^n T^{-t_i} \mathcal{U}).$$

The *topological sequence entropy of (X, T) along sequence S* is $h_{top}^S(T) = \sup_{\mathcal{U}} h_{top}^S(T, \mathcal{U})$, where supremum is taken over all finite open covers of X . If $S = \mathbb{Z}_+$, we recover standard topological entropy. In this case we omit the superscript $\mathbb{Z}_+$.

For an m.p.s. $(X, \mathcal{X}, T, \mu)$, a sequence $S \subset \mathbb{Z}_+$ and a partition $\xi \in \mathcal{P}_X$, the *sequence entropy* $h_{\mu}^S(T, \xi)$ and $h_{\mu}^S(T)$ can be defined similarly. $h_{top}^S(T)$ and $h_{\mu}^S(T)$ are conjugacy and isomorphism invariants respectively.

Let $(X, \mathcal{X}, \mu, T)$ be an m.p.s. If $h_{\mu}(T) > 0$, then $h_{\mu}^S(T) = K(S)h_{\mu}(T)$, where $K(S)$ is a number and does not dependent on T (Krug and Newton 1972). This result implies that sequence entropy is uninteresting as a new invariant in case T has positive entropy. However, little is known in case $h_{\mu}(T) = 0$.

For a t.d.s. (X, T) , Goodman (1974) showed (with a restriction that can be removed (Eberlein 1975), see also (Huang and Ye 2009)) that for any $S \in \mathcal{F}_{inf}$

$$h_{top}^S(T) \geq \sup_{\mu \in \mathcal{M}(X, T)} h_{\mu}^S(T),$$

with an equality in case $h_{top}(T) > 0$. If $h_{top}(T) = 0$, then the variational principle for topological sequence entropy needs not to hold. Goodman

gives an example with $h_{top}^S(T) = \log 2$ but $\sup_{\mu \in \mathcal{M}(X, T)} h_{\mu}^S(T) = 0$ (Goodman 1974).

Note that for an m.p.s. $(X, \mathcal{X}, \mu, T)$, if $h_{\mu}(T) > 0$, then for all $S \in \mathcal{F}_{inf}$, $h_{\mu}^S(T) > 0$ (Saleski 1977). Similarly for a t.d.s. (X, T) , if $h_{top}(T) > 0$, then for all $S \in \mathcal{F}_{inf}$, $h_{top}^S(T) > 0$ (Huang et al. 2005). If X is a countable compact metric space, it is known that $h_{top}(T) = 0$ for any t.d.s. (X, T) , and this is not the case for the topological sequence entropy (Ye and Zhang 2008).

Similar to the entropy tuples, one can define sequence entropy n -tuple and sequence entropy n -tuple for μ . They have lots of similar properties with entropy tuples. But since there is no variational principle for sequence entropy, we do not have local variational principles for sequence entropy tuples. We will not discuss sequence entropy tuples in details, and we refer to Huang et al. (2003, 2004), Huang and Ye (2009), Kerr and Li (2007), and Maass and Shao (2007), etc. to interested readers.

It is found that to characterize different mixing properties using sequence entropy related to some sequences is effective and fruitful. Here we just give an example to make the point. For more about this topic, we refer to Coronel et al. (2009), Hulse (1982, 1986), Huang et al. (2005), Saleski (1977), and Zhang (1992, 1993).

Theorem 3.9 (Huang et al. 2005) *Let $(X, \mathcal{X}, \mu, T)$ be an m.p.s. Then the following statements are equivalent:*

1. $(X, \mathcal{X}, \mu, T)$ is mildly mixing;
2. For any $\alpha \in \mathcal{P}_X$ by two nontrivial elements and IP-set F , there exists an infinite sequence $A \subseteq F$ such that $h_{\mu}^A(T, \alpha) > 0$;
3. For any $\alpha \in \mathcal{P}_X$ by finite nontrivial elements and IP-set F , there exists an infinite sequence $A \subseteq F$ such that $h_{\mu}^A(T, \alpha) > 0$.

In the above theorem a finite measurable set $C \in \mathcal{X}$ is nontrivial if $0 < \mu(C) < 1$. We have similar result for weakly mixing systems. In the topological case we have:

Theorem 3.10 (Huang et al. 2005) *Let (X, T) be a t.d.s. Then the following statements are equivalent:*

1. (X, T) is topologically mildly mixing.
2. For any cover $\mathcal{U}$ of X by two non-dense open sets and IP-set F , there exists an infinite sequence $A \subseteq F$ such that $h_{top}^A(T, \mathcal{U}) > 0$.
3. For any cover $\mathcal{U}$ of X by finite non-dense open sets and IP-set F , there exists an infinite sequence $A \subseteq F$ such that $h_{top}^A(T, \mathcal{U}) > 0$.

We also have a similar result for topologically weakly mixing systems.

Null Systems: Measurable Versus Topological

An m.p.s. $(X, \mathcal{X}, \mu, T)$ is called *null*, if $h_\mu^S(T) = 0$ for any $S \in \mathcal{F}_{inf}$. The following Kushnirenko's theorem gives the characterization of discrete spectrum via sequence entropy.

Theorem 3.11 (Kushnirenko 1967) *An m.p.s. $(X, \mathcal{X}, \mu, T)$ has discrete spectrum if and only if it is null.*

In fact, for an m.p.s. $(X, \mathcal{X}, \mu, T)$ and $\alpha \in \mathcal{P}_X$

$$\max_{S \in \mathcal{F}_{inf}} h_\mu^S(T, \alpha) = H_\mu(\alpha | \mathcal{K}_\mu),$$

where $\mathcal{K}_\mu$ is the Kronecker algebra (Huang et al. 2004). As shown in Huang et al. (2004, 2005), for $B \in \mathcal{B}$ and $R \in \mathcal{F}_{inf}$, $\text{cl}(\{\bigcup^n 1_B : n \in R\})$ is a compact set of $L^2(\mu)$ if and only if $h_\mu^S(T, \{B, X \setminus B\}) = 0$ for each infinite sequence $S \subseteq R$. In particular, $B \in \mathcal{K}_\mu$ if and only if $h_\mu^S(T, \{B, X \setminus B\}) = 0$ for any $S \in \mathcal{F}_{inf}$.

A t.d.s. (X, T) is null if $h_{top}^S(T) = 0$ for any $S \in \mathcal{F}_{inf}$. Note that it is easy to show that an equicontinuous system is null. It is natural to conjecture that if a minimal system is null then it is also equicontinuous. Unfortunately this is not the case, see for example (Goodman 1974). But we have that for a minimal system Kushnirenko's statement remains true modulo an almost one-to-one extension. Note that $\pi : (X, T) \rightarrow (Y, S)$ is almost one to one if $\{x \in X : |\pi^{-1}(\pi(x))| = 1\}$ is a dense G_δ subset.

Theorem 3.12 (Huang et al. 2003) *If a minimal t.d.s. (X, T) is null, then it is an almost one-to-one extension of its maximal equicontinuous factor (X_{eq}, T_{eq}) . Moreover, it is uniquely ergodic and has a discrete spectrum with respect to the unique measure.*

Recently, this result was improved by Fuhrmann et al. (2018). They showed that if a minimal t.d.s. (X, T) is null, then it is a regular extension of (X_{eq}, T_{eq}) , that is, if $\pi : (X, T) \rightarrow (X_{eq}, T_{eq})$ is the factor map and $\mu \in \mathcal{M}(X, T)$ be the unique measure on X , then $\mu(\{x \in X : |\pi^{-1}(\pi(x))| = 1\}) = 1$. Note that for a minimal system, whether $\pi : X \rightarrow X_{eq}$ is regular can be interpreted by the so-called diam mean equicontinuity (García-Ramos et al. 2019).

Note that the nullness in non-minimal systems was studied in Qiu and Zhao (n.d.). Moreover, the so-called tame system was extensively studied, see (Fuhrmann et al. 2018; Glasner 2018; Huang 2006; Kerr and Li 2007). Note that a minimal null systems is tame (Kerr and Li 2007).

It is an open problem (see (Huang et al. 2003)): Does a transitive non-minimal null system exist?

Maximal Pattern Entropy

The notion of maximal pattern entropy. was introduced in Huang and Ye (2009). For a t.d.s. (X, T) , $n \in \mathbb{N}$ and a finite open cover $\mathcal{U}$ let

$$p_{X, \mathcal{U}}^*(n) = \max_{(t_1 < t_2 < \dots < t_n) \in \mathbb{Z}_+^n} N(\bigvee_{i=1}^n T^{-t_i} \mathcal{U}).$$

The maximal pattern entropy of T with respect to $\mathcal{U}$ is defined by

$$h_{top}^*(T, \mathcal{U}) = \lim_{n \rightarrow +\infty} \frac{1}{n} \log p_{X, \mathcal{U}}^*(n).$$

The maximal pattern entropy of (X, T) is $h_{top}^*(T) = \sup_{\mathcal{U}} h_{top}^*(T, \mathcal{U})$, where supremum is taken over all finite open covers of X .

Analogously, given an m.p.s. $(X, \mathcal{X}, \mu, T)$ we can define the maximal pattern entropy $h_\mu^*(T)$. We remark that one of equivalent definitions is that the maximal pattern entropy is the supremum of sequence entropies, that is,

Theorem 3.13 (Huang and Ye 2009) *For a t.d.s. (X, T) we have that $h_{top}^*(T) = \sup_{S \in \mathcal{F}_{inf}} h_{top}^S(T)$ and for an m.p.s. $(X, \mathcal{X}, \mu, T)$ we have $h_\mu^*(T) = \sup_{S \in \mathcal{F}_{inf}} h_\mu^S(T)$.*

Thus a t.d.s. (X, T) is null if and only if $h_{top}^*(T) = 0$; an m.p.s. $(X, \mathcal{X}, \mu, T)$ has discrete spectrum if and only if $h_\mu^*(T) = 0$. The maximal pattern entropy has some interesting properties. For example, they are conjugacy or isomorphism invariant. Moreover, $k \in \mathbb{N} \setminus \{0\}$, $h_{top}^*(T^k) = h_{top}^*(T)$, and $h_\mu^*(T^k) = h_\mu^*(T)$. An interesting result about the value of the maximal pattern entropy is as follows:

Theorem 3.14

1. For a t.d.s. (X, T) , $h_{top}^*(T) \in \{\log k : k \in \mathbb{N}\} \cup \{\infty\}$ (Huang and Ye 2009).
2. For an ergodic m.p.s. $(X, \mathcal{X}, \mu, T)$, $h_\mu^*(T) \in \{\log k : k \in \mathbb{N}\} \cup \{\infty\}$ (Saleski 1977).

In fact, by introducing sequence entropy tuples, one can show that for a t.d.s. (X, T) , $h_{top}^*(T)$ is $\log k$ for some $k \in \mathbb{N} \cup \{\infty\}$ with k the maximal length of intrinsic sequence entropy tuples. For an m.p.s. $(X, \mathcal{X}, \mu, T)$, if $h_\mu^*(T) > \log(n-1)$ for $n \geq 2$, then there is an intrinsic μ -sequence entropy tuple of length n ; conversely, if there is an intrinsic μ -sequence entropy tuple of length n and μ is ergodic, then $h_\mu^*(T) \geq \log n$. See (Huang and Ye 2009) for details.

Theorem 3.12 says that if a minimal system (X, T) is null, that is, $h_{top}^*(T) = 0$ then (X, T) is an almost one-to-one extension of its maximal equicontinuous factor and is uniquely ergodic. There is a natural question: if a minimal system (X, T) is bounded, that is, $h_{top}^*(T) < \infty$, then what will happen? Recently Huang et al. showed that each bounded minimal system is an almost finite to one extension of its maximal equicontinuous factor, and it has only finitely many ergodic measures (Huang et al. 2019b).

For a compact metric space X , let $S(X)$ be the set of the values $h_{top}^*(T)$ for all continuous self-maps T on X . It is clear that $\{0\} \subseteq S(X) \subseteq \{0, \log 2, \log 3, \dots\} \cup \{\infty\}$ by Theorem 3.14. It was shown

in Cánovas (2004), Tan et al. (2010), and Tan (2011) that $S(X) = \{0, \log 2, \infty\}$ when X is a finite graph. Moreover, it was showed that $S(X) = \{0, \log 2, \log 3, \dots\} \cup \{\infty\}$ when X is a zerodimensional space with infinite derived sets (Tan et al. 2010). Recently, Snoha et al. (2018) obtained the following:

Theorem 3.15 *For every set $\{0\} \subseteq A \subseteq \{0, \log 2, \log 3, \dots\} \cup \{\infty\}$ there exists a one-dimensional continuum $X_A \subseteq \mathbb{R}^3$ with $S(X_A) = A$.*

For the related work, see (Cánovas 2004; Tan et al. 2010; Tan 2011).

Other Tuples

We want to mention that the idea to study the local properties of tuples can be applied in many other situations. For example, the notion of weakly mixing tuples can be defined and used in the study of weakly mixing systems and null systems (Huang et al. 2003). Tame systems are an important class of zero entropy systems. The notion of IT -tuples is a main tool to study the properties (Huang 2006; Kerr and Li 2007). The notion of sensitive tuples is also a useful tool to study sensitivity both in topological dynamics and ergodic theory (Huang et al. 2011a). The notion of $\mathcal{F}_{fp}$ -pairs and its application can be found in Dong et al. (2013). Moreover, the notion of entropy tuples can be generalized to entropy points, sequence, and sets (Dou et al. 2006; Ye and Zhang 2007).

We believe that such ideas can be applied to more situations.

Structure Theorems and Multiple Ergodic Averages

To answer the question if a minimal distal t.d.s. is an equicontinuous one, Furstenberg obtained the structure theorem for a minimal distal system. Then the structure theorem for the general minimal t.d.s. was built. Almost at the same time, Furstenberg and Zimmer obtained the structure theorem for an ergodic m.p.s. It should be noted that the theorem has an important role to get an

ergodic proof of the well-known Szemerédi's theorem. To solve the problem if the multiple ergodic averages converge in L^2 -norm, Host-Kra obtained a structure theorem of an ergodic m.p.s. involving nilsystems. The corresponding structure theorem of a minimal t.d.s. involving nilsystems was built by Host-Kra-Maass and Shao-Ye. Here we will review those results. Moreover, we will show how topological methods can be used to prove the pointwise converges of multiple ergodic averages for distal systems.

Structure Theorems in Topological Dynamics

The structure theory of minimal systems, originated in Furstenberg's seminal work (Furstenberg 1963). It was mainly developed for group actions on Hausdorff spaces. Here we assume for the rest of the paper that T is a homeomorphism, and we will also restrict to metrizable systems. We refer the reader to Auslander (1988), Bronstein (1979), Ellis (1969), Glasner (1976), Veech (1977), and de Vries (1993) for details.

We first recall definitions of extensions. Let $\pi : (X, T) \rightarrow (Y, S)$ be an extension of t.d.s. Write

$$R_\pi = \{(x_1, x_2) \in X^2 : \pi(x_1) = \pi(x_2)\}.$$

We say that π is *proximal* if $R_\pi \subset P(X, T)$, where $P(X, T)$ is the set of all proximal pairs. An extension π is an *equicontinuous extension* if for every $\varepsilon > 0$ there is $\delta > 0$ such that $(x, y) \in R_\pi$ and $\rho(x, y) < \delta$ imply $\rho(T^n x, T^n y) < \varepsilon$, for every $n \in \mathbb{Z}$. An equicontinuous extension is also called an *isometric* or *almost periodic* extension. An extension π is a *(topologically) weakly mixing extension* if $(R_\pi, T \times T)$ as a subsystem of the product system $(X \times X, T \times T)$ is transitive.

Definition 14 We say that a minimal system (X, T) is a *strictly PI system* (PI means proximal-isometric) if there is an ordinal η (which is countable since X is metrizable) and a family of systems $\{(W_\iota, w_\iota)\}_{\iota \leq \eta}$ such that

1. W_0 is the trivial system,

2. for every $\iota < \eta$ there exists an extension $\phi_\iota : W_{\iota+1} \rightarrow W_\iota$ which is either proximal or equicontinuous,
3. for a limit ordinal $\nu < \eta$ the system W_ν is the inverse limit of the systems $\{W_\iota\}_{\iota < \nu}$,

$$W_\eta = X.$$

We say that (X, T) is a *PI-system* if there exists a strictly PI system $\tilde{X}$ and a proximal extension $\theta : \tilde{X} \rightarrow X$.

If in the definition of PI-systems, we replace proximal extensions by almost one-to-one extensions we get the notion of *HPI-systems*. If we replace the proximal extensions by trivial extensions (i.e., we do not allow proximal extensions at all), we have *I-systems*.

In this terminology Furstenberg's structure theorem for a distal system and the Veech-Ellis structure theorem for a point distal system can be stated as follows:

Theorem 4.1 (Furstenberg 1963, Furstenberg) *A minimal t.d.s. is distal if and only if it is an I-system.*

Theorem 4.2 (Veech 1970; Ellis 1973, Veech-Ellis) *A minimal t.d.s. is point distal if and only if it is an HPI-system.*

Finally we state the structure theorem for the general minimal systems, which was built in Ellis et al. (1975), McMahon (1976), Veech (1977), and Glasner (1976). Roughly speaking, the class of minimal flows is the smallest class of flows containing the trivial flow and closed under (a) homomorphisms, (b) inverse limits, and (c) three “building blocks” which are equicontinuous extensions, proximal extensions, and topologically weakly mixing extensions. To be precise, we have the following structure theorem for minimal systems.

Theorem 4.3 (Structure theorem for minimal systems). *Let (X, T) be a minimal t.d.s.*

Then we have the following diagram:

$$\begin{array}{c} X \leftarrow X_\infty \\ \downarrow \pi_\infty \\ Y_\infty \end{array}$$

where X_∞ is a proximal extension of X and an RIC topologically weakly mixing extension of the strictly PI-system Y_∞ . The extension π_∞ is an isomorphism (so that $X_\infty = Y_\infty$ if and only if X is a PI-system).

In the theorem, that the extension π_∞ is a *relatively incontractible (RIC) extension* means that it is open and for every $n \geq 1$ the minimal points are dense in the relation

$$R_\pi^n = \{(x_1, \dots, x_n) \in X^n : \pi(x_j), \forall 1 \leq i \leq j \leq n\}.$$

Note that $R_\pi^1 = X$ and $R_\pi^2 = R_\pi$. For applications of the structure theorem for minimal systems, we refer (Akin et al. 2010; Auslander 1988; Bronstein 1979; Ellis 1969; Glasner 1994, 1996, 2003; Shao and Ye 2012); etc.

Structure Theorems in Ergodic Theory

Inspired by Furstenberg's structure theorem for distal systems, Parry in 1967 suggested an intrinsic definition of measurable distality (Parry 1967). In Zimmer (1976a, b), Zimmer developed the theory of distal systems (for a general locally compact acting group) and lead to a structure theorem for the general ergodic system. Independently, and at about the same time, Furstenberg proved the same theorem, which was used as the main tool for his proof of Szemerédi theorem on arithmetical progressions (Furstenberg 1977, 1981). Roughly speaking, Furstenberg-Zimmer structure theorem says that the class of ergodic systems is the smallest class of flows containing the trivial flow and closed under (a) homomorphisms, (b) inverse limits, and (c) two “building blocks” which are isometric extensions and weakly mixing extensions.

First we give the definition of measurable distality, and then we state the Furstenberg-Zimmer structure theorem.

Definition 15 (Parry 1967) Let (X, X, μ, T) be an ergodic m.p.s. A sequence $A_1 \supset A_2 \supset \dots$ of sets in X with $\mu(A_n) > 0$ and $\mu(A_n) \rightarrow 0, n \rightarrow \infty$, is called a *separating sieve* if there exists a subset $X_0 \in X$ with $\mu(X_0) = 1$ such that for every $x; x' \in X_0$ the condition “for every $n \in \mathbb{N}$ there exists $k \in \mathbb{Z}$ with $T^k x, T^k x' \in A_n$ ” implies $x = x'$. We say that an ergodic m.p.s. (X, X, μ, T) is *measurable distal* if either X is finite or there exists a separating sieve.

Parry (1967) showed that if a minimal t.d.s. (X, T) is distal, then for any $\mu \in \mathcal{M}(X, T)$, $(X, \mathcal{B}_X, \mu, T)$ is measurable distal. Also any measurable distal m.p.s. has zero entropy, and hence by the variational principle any distal t.d.s. has zero topological entropy.

Let $\pi : (X, X, \mu, T) \rightarrow (X, \mathcal{Y}, \nu, T)$ be a factor map between two m.p.s. and $\mu = \int_Y \mu_y d\nu(y)$ the disintegration of μ relative to ν . A function $f \in L^2(X, X, \mu)$ is *almost periodic over $\mathcal{Y}$* if for every $\varepsilon > 0$ there exist $g_1, \dots, g_l \in L^2(X, X, \mu)$ such that for all $n \in \mathbb{Z}$

$$\min_{1 \leq j \leq l} \|T^n f - g_j\|_{L^2(\mu_y)} < \varepsilon$$

for ν almost every $y \in Y$. Let $K(X|Y, T)$ be the closed subspace of $L^2(X)$ spanned by the almost periodic functions over $\mathcal{Y}$. When $\mathcal{Y}$ is trivial, $K(X|Y, T)$ is $H_c = L^2(X, \mathcal{K}_\mu, \mu)$, the closed subspace spanned by eigenfunctions of T .

We say that X is an *isometric extensions* of Y if $K(X|Y, T) = L^2(X)$ and it is a *(relatively) weak mixing extension* of Y if $K(X|Y, T) = L^2(Y)$.

An ergodic system X is an isometric extension of (Y, ν, S) if and only if X is isomorphic to a *skew product* $X' = Y \times M$, where $M = G/H$, G is a compact metric space, H a closed subgroup and $\mu' = \nu \times m_M$ with m_M is the Haar measure of M , that is the unique probability measure on M invariant under the action of G on M by left translations. Moreover, the action of T' on X' is given by

$$T'(y, gH) = (Sy, \alpha(y)gH),$$

where $\alpha : Y \rightarrow G$ is a measurable map, called the *cocycle* defining the extension. We denote X' by $Y \times_{\alpha} G/H$, and T' by T_{α} .

Theorem 4.4 (Zimmer). (Zimmer 1976b) *Let (X, X, μ, T) be an ergodic m.p.s. Then (X, X, μ, T) is measurable distal if and only if there exists a countable ordinal η and a directed family of factors $(X_{\theta}, X_{\theta}, \mu_{\theta}, T)$, $\theta \leq \eta$ such that*

1. $X_0 = \{\text{pt}\}$ is the trivial system and $X_{\eta} = X$.
2. For $\theta < \eta$ the extension $\pi_{\theta} : X_{\theta+1} \rightarrow X_{\theta}$ is isometric and nontrivial (i.e., not an isomorphism).
3. For a limit ordinal $\lambda \leq \eta$, $X_{\lambda} = \lim_{\leftarrow \theta < \lambda} X_{\theta}$ (i.e. $X_{\lambda} = \bigvee X_{\theta}$).

Here is the Furstenberg-Zimmer structure theorem for the action $\mathbb{Z}$. For the general locally compact acting group actions, we refer to Furstenberg (1981), Glasner (2003), Zimmer (1976a, b).

Theorem 4.5 (Furstenberg-Zimmer's structure theorem). *Each ergodic m.p.s. (X, X, μ, T) is a weakly mixing extension of an ergodic distal system X_{η} .*

$$\underbrace{X_0 \xleftarrow{\pi_0} X_1 \xleftarrow{\pi_1} \cdots \xleftarrow{\pi_{n-1}} X_n \xleftarrow{\pi_n} X_{n+1} \xleftarrow{\pi_{n+1}} \cdots \xleftarrow{\pi_{\eta}} X_{\eta} \xleftarrow{\pi_{\eta}} X}_{\text{distalfactor}}.$$

Structure theorems are very useful in the study of combinatorial number theory, we refer to Bergelson (2006), Bergelson and Leibman (1996), Furstenberg (1977, 1981, 1991), Frantzikinakis (2017), and Kra (2006), etc. for more details.

Here we mention an example of the applications of the structure theorems in the theory of chaos. The notion of Li-Yorke chaos was introduced in Li and Yorke (1975). Let (X, T) be a t.d.s. A pair $(x, y) \in X \times X$ is called a *Li-Yorke pair* if

$$\liminf_{n \rightarrow +\infty} \rho(T^n x, T^n y) = 0 \text{ and } \limsup_{n \rightarrow +\infty} \rho(T^n x, T^n y) > 0.$$

A subset C of X is called *scrambled* if any two distinct points $x, y \in C$ form a scrambled pair. A t.d.s. (X, T) is said to be *Li-Yorke chaotic* if there is an uncountable scrambled set in X . Using Furstenberg-Zimmer structure theorem, Blanchard et al. (2002a) showed that positive entropy implies Li-Yorke chaos. In fact, they showed that if for a t.d.s. (X, T) there is some ergodic measure μ such that $(X, \mathcal{B}_X, \mu, T)$ is not measurable distal, then (X, T) is Li-Yorke chaotic. Note that Kerr and Li gave an alternative proof of this fact using the combinatorial argument (Kerr and Li 2007). Using the structure of minimal systems, Akin et al. showed that if a minimal system (X, T) is not PI, then it is Li-Yorke chaotic (Akin et al. 2010). It is known that if each pair not in the diagonal is (positively) Li-Yorke, then it has zero entropy (Blanchard et al. 2002b). It is an open question if this is still true if we replace positively Li-Yorke by positively or negatively Li-Yorke.

We refer the readers to Blanchard and Huang (2008), Downarowicz (2014), Huang (2008), Huang and Jin (2016), Huang et al. (2014, 2015, 2017a), Huang and Lu (2017), Li and Qiao (2018), Wang et al. (2019), Zhang (2006), Zhou et al. (2017), and Liu (2019) for more study on chaotic phenomenon appearing in positive entropy systems by using the variational principle and the structure theorems.

Multiple Ergodic Averages

Ramsey theory is best defined by example and the classic example of a Ramsey type theorem is the result of van der Waerden (1927) in the 1920s: if the integers are partitioned into finitely many subsets, at least one of the subsets contains arbitrarily long arithmetic progressions. Erdős and Turán (1936) conjectured that a weaker assumption suffices: if E is a set of integers whose upper density $\bar{d}(E)$ is positive, then E contains arbitrarily long arithmetic progressions. This conjecture was verified by Szemerédi in 1975 (Szemerédi 1975), and was reproved by Furstenberg using ergodic theoretic methods in 1976 (Furstenberg 1977). Green and Tao showed that the primes contain arbitrarily long arithmetic progressions (Green and Tao

2008). Now there are lots of excellent surveys concerning this topic, see for instance (Bergelson 2006; Frantzikinakis and McCutcheon 2012; Furstenberg 1988, 1991, 2010; Kra 2006; Ward 2012), etc. The recent book by Host and Kra (2018) is an excellent source for these topics.

To prove Szemerédi theorem, Furstenberg (1977) showed that if $(X, \mathcal{X}, \mu, T)$ is a m.p.s., $f \in L^\infty(X, \mathcal{X}, \mu)$, $f \geq 0$ and not identically 0, then for any k ,

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int f(x) f(T^n x) \dots f(T^{(k-1)n} x) d\mu(x) > 0. \quad (1)$$

As $A_N(f, x) = \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)$ is called the ergodic average, the averages above are called *multiple ergodic averages* (or “nonconventional averages” (Furstenberg 1988)). More generally, we consider the averages

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T_1^{p_1(n)} x) f_2(T_2^{p_2(n)} x) \dots f_d(T_d^{p_d(n)} x) \quad (2)$$

where $T_1, T_2, \dots, T_d$ are invertible measure preserving transformations of a probability space $(X, \mathcal{X}, \mu)$, and $p_1, \dots, p_d$ are integer valued polynomials. The need of a polynomial character is not artificial, it is meaningful for combinatorial and diophantine applications (Bergelson 2006; Furstenberg 2010).

The basic question related to the multiple ergodic averages is as follows:

Question 4.6 *Let $T_1, T_2, \dots, T_d$ be measure preserving transformations of a probability space $(X, \mathcal{X}, \mu)$ generating a nilpotent group (of transformations). Is it true that for any integer valued polynomials $p_i(n)$ and $f_i \in L^\infty(X, \mathcal{X}, \mu)$, $i = 1, 2, \dots, d$,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f_1(T_1^{p_1(n)} x) f_2(T_2^{p_2(n)} x) \dots f_d(T_d^{p_d(n)} x) \quad (3)$$

exists in the $L^2(\mu)$ norm? Almost everywhere?

It was shown in Bergelson and Leibman (2002) that this multiple ergodic average may not converge in general when the group generated by $T_1, \dots, T_d$ is only assumed to be solvable.

For an m.p.s. $(X, \mathcal{X}, \mu, T)$, there were numerous partial results, including the works of Conze and Lesigne (1984, 1987, 1988) for the average of an arithmetic progression with three terms, the work of Furstenberg and Weiss (1996) for a polynomial average with two terms, as well as other partial convergence results (Host and Kra 2001; Host and Kra 2002). Mean convergence for arithmetic progressions was proven by Host and Kra (2005) (for another proof, see (Ziegler 2007)), and their proof motivated the new structural description of an m.p.s. We will discuss their proof later.

Walsh (2012) proved the convergence for the multiple ergodic averages where the transformations are taken from a nilpotent group of transformations and the exponents are arbitrary integer valued polynomials.

Theorem 4.7 (Walsh). (Walsh 2012) *Let G be a nilpotent group of measure preserving transformations of a probability space $(X, \mathcal{X}, \mu)$. Then, for every $T_1, \dots, T_d \in G$, the averages*

$$\frac{1}{N} \sum_{n=1}^N \prod_{j=1}^l (T_1^{p_{1,j}(n)} \dots T_d^{p_{d,j}(n)}) f_j$$

always converge in $L^2(X, \mathcal{X}, \mu)$ for every $f_1, \dots, f_d \in L^\infty(X, \mathcal{X}, \mu)$ and every set of integer valued polynomials $p_{i,j}$.

Walsh’s proof is not ergodic in nature, but rather combinatorial, following previous work of Tao (2008). This new method sufficed to prove convergence results, but it lacked the structural description of the averages, the associated description of the limit, and the combinatorial theorems that can be derived from this structure. As mentioned in Host and Kra (2018), an

interesting, but very difficult, problem is to find a way to give structural results that control these more general averages. Such results are not yet available in the general setting of nilpotent groups of transformations, nor even just for commuting transformations. We refer to Furstenberg (2010), Kra (2006), and Host and Kra (2018) for more details on this topic.

Characteristic Factors in Ergodic Theory

In the study of multiple ergodic averages, the idea of characteristic factors plays a very important role. This idea was suggested by Furstenberg in Furstenberg (1977), and the notion of “characteristic factors” was first introduced in a paper by Furstenberg and Weiss (1996). Let $(X, \mathcal{X}, \mu, T)$ be an m.p.s. and $(Y, \mathcal{Y}, \nu, S)$ be a factor of X . Let $\{p_1, \dots, p_d\}$ be a family of integer valued polynomials, $d \in \mathbb{N}$. We say that Y is a *characteristic factor* of X for the scheme $\{p_1, \dots, p_d\}$ if for all $f_1, \dots, f_d \in L^\infty(X, \mathcal{X}, \mu)$,

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} T^{p_1(n)} f_1 T^{p_2(n)} f_2 \dots T^{p_d(n)} f_d - \frac{1}{N} \sum_{n=0}^{N-1} T^{p_1(n)} \mathbb{E}(f_1 | \mathcal{Y}) T^{p_2(n)} \mathbb{E}(f_2 | \mathcal{Y}) \dots T^{p_d(n)} \mathbb{E}(f_d | \mathcal{Y}) \right\|_{L^2} \rightarrow 0.$$

Finding a characteristic factor often gives a reduction of the problem of evaluating limit behavior of multiple averages to special systems. In the rest of the entry, we mainly consider the scheme $\{n, 2n, \dots, dn\}$. That is, we consider the following multiple ergodic averages:

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) f_2(T^{2n} x) \dots f_d(T^{dn} x). \quad (4)$$

To study (4), it suffices to consider ergodic systems and that we will restrict to this case below. Furstenberg (1977) proved that the distal factor is a characteristic factor for (4), and this observation was a main step in the proof of the ergodic version of Szemerédi theorem, see (1). But this fact is not enough to show the mean convergence of (4). It always exists a minimal

characteristic factor Z , meaning that Z is characteristic and that every proper factor of Z is not characteristic. To describe this minimal characteristic factor, we need to introduce some notions.

Definition 16 Let G be a group. For $A, B \subset G$, we write $[A, B]$ for the subgroup spanned by $\{[a, b] = aba^{-1}b^{-1} : a \in A, b \in B\}$. The commutator subgroups $G_j, j \geq 1$, are defined inductively by setting $G_1 = G$ and $G_{j+1} = [G_j, G]$. Let $d \geq 1$ be an integer. We say that G is *d-step nilpotent* if G_{d+1} is the trivial subgroup.

Let G be a d -step nilpotent Lie group and Γ be a discrete cocompact subgroup of G . The compact manifold $X = G/\Gamma$ is called a *d-step nilmanifold*. The group G acts on X by left translations and we write this action as $(g, x) \mapsto gx$. The Haar measure μ of X is the unique probability measure on X invariant under this action. Let $\tau \in G$ and T be the transformation $x \mapsto \tau x$ of X . Then (X, μ, T) is called a *d-step nilsystem*.

Conze and Lesigne had shown (Conze and Lesigne 1984, 1987, 1988) that an inverse limit of 2-step nilsystems is the characteristic factor for the 3-term multiple ergodic averages. The general case was confirmed by constructing such factors in Host and Kra (2005) (see also (Ziegler 2007)).

Now we outline Host-Kra’s methods and results, and we need some notations first. Let X be a set, and let $d \geq 1$ be an integer. We view element in $\{0, 1\}^d$ as a sequence $\varepsilon = \varepsilon_1 \dots \varepsilon_d$ of 0’s and 1’s, and let $|\varepsilon| = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_d$. We denote X^{2^d} by $X^{[d]}$. A point $x \in X^{[d]}$ can be written as $x = (x_\varepsilon : \varepsilon \in \{0, 1\}^d)$. A point $x \in X^{[d]}$ can be decomposed as $x = (x', x'')$ with $x', x'' \in X^{[d-1]}$, where $x' = (x_{\varepsilon 0} : \varepsilon \in \{0, 1\}^{d-1})$ and $x'' = (x_{\varepsilon 1} : \varepsilon \in \{0, 1\}^{d-1})$, which induces a natural identification of $X^{[d]}$ and $X^{[d-1]} \times X^{[d-1]}$. As examples, points in $X^{[2]}$ are like $(x_{00}, x_{01}, x_{10}, x_{11})$.

Let $(X_1, \mathcal{X}_1, \mu_1, T_1), (X_2, \mathcal{X}_2, \mu_2, T_2)$ be two m.p.s. and let $(Y, \mathcal{Y}, \nu, S)$ be a common factor with $\pi_i : X_i \rightarrow Y$ for $i = 1, 2$ the factor maps. Let $\mu_i = \int \mu_{i,y} d\nu(y)$ represent the disintegration of μ_i with respect to Y . Let $\mu_1 \times_{\mathcal{Y}} \mu_2$ denote the measure defined by

$$\mu_1 \times_Y \mu_2(A) = \int_Y \mu_{1,y} \times \mu_{2,y} dv(y),$$

for all $A \in X_1 \times X_2$. The system $(X_1 \times X_2, X_1 \times X_2, \mu_1 \times_Y \mu_2, T_1 \times T_2)$ is called the *relative product* of X_1 and X_2 with respect to Y and is denoted $X_1 \times_Y X_2$. $\mu_1 \times_Y \mu_2$ is also called *relatively independent joining* of X_1 and X_2 over Y .

For an m.p.s. (X, X, μ, T) we write $I(T)$ for the σ -algebra $\{A \in X : T^{-1}A = A\}$ of invariant sets. Let (X, X, μ, T) be an ergodic system and $k \in \mathbb{N}$. We define a measure $\mu^{[k]}$ on $X^{[k]}$ invariant under $T^{[k]} = T \times T \times \dots \times T$ (2^k times), by $\mu^{[1]} = \mu \times_{I(T)} \mu = \mu \times \mu$; and for $k \geq 1$,

$$\mu^{[k+1]} = \mu^{[k]} \times_{I(T^{[k]})} \mu^{[k]}.$$

For an integer $k \geq 1$, let (Ω_k, P_k) be the system corresponding to the σ -algebra $I^{[k]}$ and let

$$\mu^{[k]} = \int_{\Omega_k} \mu_{\omega}^{[k]} dP_k(\omega)$$

denote the ergodic decomposition of $\mu^{[k]}$ under $T^{[k]}$. Then by definition

$$\mu^{[k+1]} = \int_{\Omega_k} \mu_{\omega}^{[k]} \times \mu_{\omega}^{[k]} dP_k(\omega).$$

If (X, μ, T) is weakly mixing, then by induction $I(T^{[k]})$ is trivial and $\mu^{[k]}$ is the 2^k Cartesian power μ^{2^k} of μ for $k \geq 1$.

The measure $\mu^{[k]}$ is invariant under $T^{[k]}$ and each of the 2^k natural projections of $\mu^{[k]}$ on X is equal to μ . Letting $C : \mathbb{C} \rightarrow \mathbb{C}$ denote the conjugacy map $z \mapsto \bar{z}$, we have that for a bounded function f on X , the integral

$$\int_{X^{[k]}} \prod_{\varepsilon \in \{0,1\}^k} C^{|\varepsilon|} f(x_{\varepsilon}) d\mu^{[k]}(\mathbf{x})$$

is real and nonnegative. Therefore, we can define a seminorm $\|\cdot\|_k$ on $L^{\infty}(\mu)$ by

$$\|f\|_k = \left(\int_{X^{[k]}} \prod_{\varepsilon \in \{0,1\}^k} C^{|\varepsilon|} f(x_{\varepsilon}) d\mu^{[k]}(\mathbf{x}) \right)^{1/2^k}.$$

That $\|\cdot\|_k$ is a seminorm can be proved as in Host and Kra (2005), and it is called the *Host-Kra seminorm*. The equation below follows immediately from the definition of the measures and the Ergodic Theorem, which can be considered as an alternate definition of the seminorms. For every integer $k \geq 0$ and every $f \in L^{\infty}(\mu)$, one has

$$\|f\|_{k+1} = \left(\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \|f \cdot T^n \bar{f}\|_k^{2^k} \right)^{1/2^{k+1}}.$$

Then by the Host-Kra seminorm we can define factors $(Z_{d-1}, Z_{d-1}, \mu_{d-1}, T)$.

Definition 17 Let (X, X, μ, T) be an ergodic m.p.s. For $d \in \mathbb{N}$, there exists a T -invariant σ -algebra Z_{d-1} of X such that for $f \in L^{\infty}(\mu)$,

$$\|f\|_d = 0 \text{ if and only if } \mathbb{E}(f|Z_{d-1}) = 0.$$

Let $(Z_{d-1}, Z_{d-1}, \mu_{d-1}, T)$ be the factor of X associated to the sub σ -algebra Z_{d-1} .

If $X = Z_{d-1}$, then X is called a system of order $d-1$.

Using van der Corput Lemma, Host and Kra showed that the original averages along arithmetic progressions is controlled by the seminorms.

Theorem 4.8 (Host and Kra 2005) *Let (X, X, μ, T) be an ergodic m.p.s. and $d \in \mathbb{N}$. For $f_1, \dots, f_d \in L^{\infty}(X, \mu)$ with $\|f_1\|_{\infty}, \dots, \|f_d\|_{\infty} \leq 1$, one has that*

$$\begin{aligned} \limsup_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} T^n f_1 T^{2n} f_2 \dots T^{dn} f_d \right\|_{L^2} \\ \leq \min_{1 \leq j \leq d} \left\{ \|f_j\|_d \right\} \end{aligned}$$

This theorem states that the factor Z_{d-1} is characteristic for the average (4). The bulk of the

work, and also the most technical portion, is devoted to the description of these factors. The structure theorem states the system $(Z_{d-1}, Z_{d-1}, \mu_{d-1}, T)$ is a (measure theoretic) inverse limit of $d-1$ -step nilsystems.

We isolate the first coordinate, writing $X_*^{[d]} = X^{2^{d-1}}$ and then writing a point $\mathbf{x} \in X_*^{[d]}$ as $\mathbf{x} = (x_0, \mathbf{x}_*)$, where $\mathbf{x}_* = (x_\varepsilon : \varepsilon \neq \mathbf{0}) \in X_*^{[d]}$ and $\mathbf{0} = 00 \dots 0 \in \{0, 1\}^d$.

Theorem 4.9 (Host-Kra). (Host and Kra 2005) *Let (X, X, μ, T) be an ergodic system and $d \in \mathbb{N}$. Then the following properties are equivalent:*

1. X is a system of order $d-1$, that is, $(X, X, \mu, T) = (Z_{d-1}, Z_{d-1}, \mu_{d-1}, T)$.
2. The system (X, X, μ, T) is a (measure theoretic) inverse limit of $d-1$ -step nilsystems.
3. $\|\cdot\|_d$ is a norm on $L^\infty(\mu)$, equivalently, $\|f\|_d = 0$ implies that $f = 0$.
4. There exists a measurable map $J : X_*^{[d]} \rightarrow X$ such that $x_0 = J\left(x_\varepsilon : \mathbf{0} \neq \varepsilon \in \{0, 1\}^d\right)$ for $\mu^{[d]}$ almost every $\mathbf{x} = (x_\varepsilon : \varepsilon \in \{0, 1\}^d) \in X_*^{[d]}$.

Z_0 is the trivial factor, and Z_1 is the Kronecker factor (i.e. $Z_1 = \mathcal{K}\mu$) and more generally, Z_k is a compact abelian group extension of Z_{k-1} . Furthermore, the sequence of factors is increasing

$$\{pt\} = Z_0 \leftarrow Z_1 \leftarrow \dots \leftarrow Z_n \leftarrow Z_{n+1} \leftarrow \dots \leftarrow X.$$

and if T is weakly mixing, then Z_k is the trivial factor for all k .

Convergence of the linear multiple ergodic average then follows easily from the general properties of nilmanifolds proved by Leibman (2005) (which was proved by Lesigne for connected groups earlier (Lesigne 1989)). See also (Bergelson et al. 2005; Ziegler 2005). For more details on this topic, we refer to the recent book (Host and Kra 2018).

Characteristic Factors in Topological Dynamics

There are several ways to study the counterpart of characteristic factors in a t.d.s. The first way was

given by Glasner in Glasner (1994). Let (X, T) be a t.d.s. and $d \in \mathbb{N}$. Let $\sigma_d = T \times T^2 \times \dots \times T^d$. (Y, T) is said to be an *topological characteristic factor of order d* if there exists a dense G_δ set Ω of X such that for each $x \in \Omega$ the orbit closure $L = \overline{\sigma_d(x^d, \sigma_d)}$ is $\pi \times \dots \times \pi$ (d times) saturated, where $x^d = (x, \dots, x)$ (d times) and $\pi : X \rightarrow Y$ is the corresponding factor map. That is, $(x_1, x_2, \dots, x_d) \in L$ if and only if $(x'_1, x'_2, \dots, x'_d) \in L$ whenever for all $1 \leq i \leq d$, $\pi(x_i) = \pi(x'_i)$. In Glasner (1994), it was shown that if (X, T) is a distal minimal system, then its largest d -step distal factor (in the Furstenberg's tower of a minimal distal system) is a topological characteristic factor of order d ; if (X, T) is a weakly mixing system (X, T) , then the trivial system is its topological characteristic factor. It is a deep open problem whether for a minimal distal system one can replace the largest d -step distal factor in Glasner's theorem by the maximal d -step pro-nilfactor.

The second way was given by Host, Kra, and Maass in Host et al. (2010). They obtained a topological structure theory involving nilsystems for all minimal distal systems, which can be viewed as an analog of the purely ergodic structure theory of (Host and Kra 2005) and the refinement of the Furstenberg's structure theorem for minimal distal systems. In Host et al. (2010), a certain generalization of the regionally proximal relation, namely $\mathbf{RP}^{[d]}$ (the regionally proximal relation of order d), was introduced and used to produce the maximal pro-nilfactors, which can be seen as the characteristic factor of the minimal system (X, T) . In the following we will give some details of this approach.

If $\mathbf{n} = (n_1, \dots, n_d) \in \mathbb{Z}^d$ and $\varepsilon \in \{0, 1\}^d$, we define $\mathbf{n} \cdot \varepsilon = \sum_{i=1}^d n_i \varepsilon_i$. Let (X, T) be a t.d.s. and let $d \geq 1$ be an integer. We define $\mathbf{Q}^{[d]}(X)$ to be the closure in $X^{[d]}$ of elements of the form

$$\left(T^{\mathbf{n} \cdot \varepsilon} x = T^{n_1 \varepsilon_1 + \dots + n_d \varepsilon_d} x : \varepsilon = \varepsilon_1 \varepsilon_2 \dots \varepsilon_d \in \{0, 1\}^d\right),$$

where $\mathbf{n} = (n_1, \dots, n_d) \in \mathbb{Z}^d$ and $x \in X$. When there is no ambiguity, we write $\mathbf{Q}^{[d]}$ instead of $\mathbf{Q}^{[d]}(X)$. An element of $\mathbf{Q}^{[d]}(X)$ is called a (dynamical) *parallelepiped of dimension d* . As

examples, $\mathbf{Q}^{[2]}$ is the closure in $X^{[2]} = X^4$ of the set $\{(x, T^m x, T^n x, T^{n+m} x) : x \in X, m, n \in \mathbb{Z}\}$. $\mathbf{Q}^{[d]}$ may be viewed as the topological correspondence of $\mu^{[d]}$.

Face transformations $T_j^{[d]} : X^{[d]} \rightarrow X^{[d]}$, $j = 1, \dots, d$ are defined as follows: for every $\mathbf{x} = (x_\varepsilon)_{\varepsilon \in \{0,1\}^d} \in X^{[d]}$

$$\left(T_j^{[d]} \mathbf{x}\right)_\varepsilon = \begin{cases} Tx_\varepsilon & \text{if } \varepsilon_j = 1; \\ x_\varepsilon & \text{if } \varepsilon_j = 0. \end{cases}$$

The *face group* of dimension d is the group $\mathcal{F}^{[d]}(X)$ of transformations of $X^{[d]}$ spanned by the face transformations. The *cube group* or *parallel-pipeded group* of dimension d is the group $\mathcal{G}^{[d]}(X)$ spanned by the diagonal transformation and the face transformations. We often write $\mathcal{F}^{[d]}$ and $\mathcal{G}^{[d]}$ instead of $\mathcal{F}^{[d]}(X)$ and $\mathcal{G}^{[d]}(X)$, respectively. It is easy to verify that $\mathbf{Q}^{[d]}$ is the closure in $X^{[d]}$ of $\{Sx^{[d]} : S \in \mathcal{F}^{[d]}, x \in X\}$. If x is a transitive point of X , then $\mathbf{Q}^{[d]}$ is the orbit closure of $x^{[d]}$ under the group $\mathcal{G}^{[d]}$. For $x \in X$, we write $x^{[d]} = (x, x, \dots, x) \in X^{[d]}$.

Theorem 4.10 *If (X, T) is minimal and $d \in \mathbb{N}$, then $(\mathbf{Q}^{[d]}, \mathcal{G}^{[d]})$ is minimal (Host et al. 2010). And for all $x \in X$, $(\overline{O(x^{[d]}, \mathcal{F}^{[d]})}, \mathcal{F}^{[d]})$ is minimal (Shao and Ye 2012).*

Theorem 4.10 can be viewed as a topological analogue of the following ergodic theorem.

Theorem 4.11 (Host and Kra 2005) *If $(X, \mathcal{X}, \mu, T)$ is an ergodic m.p.s. and $d \in \mathbb{N}$, then $(X^{[d]}, \mu^{[d]}, \mathcal{G}^{[d]})$ is ergodic. And (Ω_d, P_d) is ergodic under the action of the group $\mathcal{F}^{[d]}$.*

Compared with Theorem 4.9, the following structure theorem characterizes the inverse limits of nilsystems using dynamical parallelepipeds.

Theorem 4.12 (Host-Kra-Maass). (Host et al. 2010) *Assume that (X, T) is a minimal t.d.s. and let $d \geq 2$ be an integer. The following properties are equivalent:*

1. X is an (topologically) inverse limit of $(d-1)$ -step minimal nilsystems.

2. If $\mathbf{x}, \mathbf{y} \in \mathbf{Q}^{[d]}$ have $2^d - 1$ coordinates in common, then $\mathbf{x} = \mathbf{y}$.
3. If $x, y \in X$ are such that $(x, y, \dots, y) \in \mathbf{Q}^{[d]}$, then $x = y$.

A minimal system satisfying one of the equivalent properties above is called a (topological) system of order $(d-1)$ or a (topological) $(d-1)$ -step pro-nilsystem.

Note that any system of order d is isomorphic in the measure theoretic sense to a topological system of order d (Host et al. 2010). Let (X, T) be a system of order d , then the maximal measurable and topological factors of order j coincide, where $j \leq d$ (Dong et al. 2013).

Definition 18 Let (X, T) be a t.d.s. and let $d \in \mathbb{N}$. The points $x, y \in X$ are said to be *regionally proximal of order d* if for any $\delta > 0$, there exist $x', y' \in X$ and a vector $\mathbf{n} = (n_1, \dots, n_d) \in \mathbb{Z}^d$ such that $\rho(x, x') < \delta$, $\rho(y, y') < \delta$, and

$$\rho(T^{\mathbf{n} \cdot \varepsilon} x', T^{\mathbf{n} \cdot \varepsilon} y') < \delta \text{ for any } \varepsilon \in \{0, 1\}^d \setminus \{\mathbf{0}\}.$$

The set of regionally proximal pairs of order d is denoted by $\mathbf{RP}^{[d]}$ (or by $\mathbf{RP}^{[d]}(X, T)$ in case of ambiguity), and is called *the regionally proximal relation of order d* .

The above definition was introduced in Host and Maass (2007) and Host et al. (2010). The authors showed (Host et al. 2010) that if a system is minimal and distal, then $\mathbf{RP}^{[d]}$ is an equivalence relation, and a very deep result stating that $(X/\mathbf{RP}^{[d]}, T)$ is the maximal d -step pro-nilfactor of the system. Using the theory of Ellis semi-group, Shao and Ye (2012) showed that all these results in fact hold for arbitrarily minimal systems of abelian group actions. In a recent paper by Glasner et al. (2018), the same question is considered for a general group action, and similar results are proved.

Theorem 4.13 (Shao and Ye 2012; Host et al. 2010) *Let (X, T) be a minimal t.d.s. and $d \in \mathbb{N}$. Then*

1. $\mathbf{RP}^{[d]}(X)$ is an equivalence relation.

2. $(X_d = X/\mathbf{RP}^{[d]}, T)$ is the maximal d -step pro-nilfactor of (X, T) .

Note that $\mathbf{RP}^{[1]}$ is the classical regionally proximal relation and $X_1 = X_{eq}$ is the maximal equicontinuous factor of the system. Furthermore, the sequence of factors is increasing

$$\{pt\} = X_0 \leftarrow X_1 \leftarrow \cdots \leftarrow X_n \leftarrow X_{n+1} \leftarrow \cdots \leftarrow X.$$

and if (X, T) is topologically weakly mixing, then X_n is the trivial factor for all n .

There are other ways to study the counterpart of characteristic factors in a t.d.s. For example, in Cai and Shao (2019), the topological characteristic factors along cubes of minimal systems are studied; and in Glasner et al. (2019), the authors considered the regionally proximal relation of order d along arithmetic progressions.

Further development of the theory of cubes in an abstract setting, calling these structures nilspaces, were given by Host and Kra (2008), Camarena and Szegedy (Cai and Shao 2019; Szegedy 2012), Gutman et al. (2019a, b, 2020), Candela (2017a, b), etc.

Topological Methods in the Study of the Multiple Ergodic Averages

Though results on the convergence of multiple ergodic averages in L^2 norm are very rich now, there are a few ones about almost pointwise convergence. In 1989 Bourgain (1989) showed that $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^{p(n)}x)$ exists a.e. for all $p(n) \in \mathbb{Z}[n]$ and $f \in L^p(X, \mathcal{X}, \mu)$ with $p > 1$. Note that this result may fail when $p = 1$ (Buczolich and Mauldin 2010). And in 1990 Bourgain (1990) proved that $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f_1(T^{a_1 n}x) f_2(T^{a_2 n}x)$ exists a.e. for all f_1, f_2 in $L^\infty(X, \mathcal{X}, \mu)$. Recently the authors of the current paper find that one may use topological methods to study almost pointwise convergence of multiple ergodic averages.

Let (X, T) be a t.d.s. and $d \in \mathbb{N}$. Set $\tau_d = T \times \cdots \times T$ (d times) and $\sigma_d = T \times T^2 \times \cdots \times T^d$. Let $\langle \tau_d, \sigma_d \rangle$ be the group generated by τ_d, σ_d . For any $x \in X$, let $N_d(X, x) = \overline{O((x, \dots, x), \langle \tau_d, \sigma_d \rangle)}$, the orbit closure of $(x, \dots, x)$ (d times) under the action of the

group $\langle \tau_d, \sigma_d \rangle$. We remark that if (X, T) is minimal, then all $N_d(X, x)$ coincide, which will be denoted by $N_d(X)$. It was shown by Glasner (1994) that if (X, T) is minimal, then $(N_d(X), \langle \tau_d, \sigma_d \rangle)$ is minimal. Hence if $(N_d(X), \langle \tau_d, \sigma_d \rangle)$ is uniquely ergodic, then it is strictly ergodic. In fact, we can give a nice model for any ergodic m.p.s.:

Theorem 4.14 (Huang et al. 2019d) *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic m.p.s. and $d \in \mathbb{N}$. Then it has a strictly ergodic model $(\widehat{X}, \widehat{T})$ such that $(N_d(\widehat{X}), \langle \tau_d, \sigma_d \rangle)$ is strictly ergodic.*

Note that for each uniquely ergodic t.d.s. it has very nice convergence property: for a t.d.s. (X, Γ) with $\Gamma = \mathbb{Z}^d$, then (X, Γ) is uniquely ergodic if and only if for every continuous function $f \in C(X)$ the sequence of functions $\frac{1}{N^d} \sum_{\gamma \in [0, N-1]^d} f(\gamma x)$ converges uniformly to a constant function. Thus as an application of Theorem 4.14 we can show the following:

Theorem 4.15 (Huang et al. 2019d) *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic m.p.s. and $d \in \mathbb{N}$. Then for $f_1, \dots, f_d \in L^\infty(\mu)$ the averages*

$$\frac{1}{N^2} \sum_{(n, m) \in [0, N-1]^2} f_1(T^n x) f_2(T^{n+m} x) \cdots f_d(T^{n+(d-1)m} x)$$

converge to a constant $\mu - a. e.$

To prove Theorem 4.14, we found that not every strictly ergodic model is the one we need, and Jewett-Krieger Theorem is not enough for our purpose. Fortunately, we find that Weiss's Theorem is a right tool.

We say that $\widehat{\pi} : \widehat{X} \rightarrow \widehat{Y}$ is a *topological model* for a factor map $\pi : (X, \mathcal{X}, \mu, T) \rightarrow (Y, \mathcal{Y}, \nu, S)$ if $\widehat{\pi}$ is a topological factor map and there exist measure theoretical isomorphisms ϕ and ψ such that the diagram

$$\begin{array}{ccc} X & \xrightarrow{\phi} & \widehat{X} \\ \pi \downarrow \downarrow & & \widehat{\pi} \\ Y & \xrightarrow{\psi} & \widehat{Y} \end{array}$$

is commutative, that is, $\widehat{\pi}\phi = \psi\pi$. Weiss (1985) generalized the theorem of Jewett-Krieger to the relative case. Namely, he proved that

Theorem 4.16 (Weiss). (Weiss 1985) *If $\pi : (X, \mathcal{X}, \mu, T) \rightarrow (Y, \mathcal{Y}, \nu, S)$ is a factor map with $(X, \mathcal{X}, \mu, T)$ ergodic and $(\hat{Y}, \mathcal{B}_{\hat{Y}}, \hat{\nu}, \hat{S})$ is a uniquely ergodic model for $(Y, \mathcal{Y}, \nu, S)$, then there is a uniquely ergodic model $(\hat{X}, \mathcal{B}_{\hat{X}}, \hat{\mu}, \hat{T})$ for $(X, \mathcal{X}, \mu, T)$ and a factor map $\hat{\pi} : \hat{X} \rightarrow \hat{Y}$ which is a model for $\pi : X \rightarrow Y$.*

For $d \geq 3$ let $\pi_{d-2} : X \rightarrow Z_{d-2}$ be the factor map from X to its $d-2$ -step nilfactor Z_{d-2} . By the results of Host-Kra-Maass in Host et al. (2010), Z_{d-2} may be regarded as a topological system in the natural way. Using Weiss's Theorem there is a uniquely ergodic model $(\hat{X}, \hat{\mathcal{X}}, \hat{\mu}, \hat{T})$ for $(X, \mathcal{X}, \mu, T)$ and a factor map $\hat{\pi}_{d-2} : \hat{X} \rightarrow Z_{d-2}$ which is a model for $\pi_{d-2} : X \rightarrow Z_{d-2}$. We then show that $(\hat{X}, \hat{T})$ is what we need, that is, $(N_d(\hat{X}), \langle \tau_d, \sigma_d \rangle)$ is uniquely ergodic.

Using some results developed when proving Theorem 4.15, one has

Theorem 4.17 (Huang et al. 2019d) *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic distal system, and $d \in \mathbb{N}$. Then for all $f_1, \dots, f_d \in L^\infty(\mu)$*

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) \dots f_d(T^{dn} x)$$

converge μ a.e.

Note that by Furstenberg-Zimmer's structure theorem and Theorem 4.17 the open question (if (4) converges pointwisely) is reduced to deal with the weakly mixing extensions. Even for weakly mixing systems, the question on the pointwise convergence of (4) still remains open. A partial answer to this question was obtained by Assani (1998), who showed that if $(X, \mathcal{X}, \mu, T)$ is a weakly mixing system such that the restriction of X to its Pinsker algebra has spectral type singular w.r.t. the Lebesgue measure, then the limit of (4) exists a.e. Also in Derrien and Lesigne (1996), it was shown that the limit of (4) exists a.e. for a K-system.

Now we describe some results for a class of weakly mixing m.p.s. obtained in Gutman et al.

(2018). A joining λ of $(X_i, \mathcal{X}_i, \mu_i, T_i)$, $1 \leq i \leq d$, is pairwise independent if its projection on $X_i \times X_j$ is equal to $\mu_i \times \mu_j$ for all $i \neq j \in \{1, 2, \dots, d\}$, and it is independent if it is the product measure. A system $(X, \mathcal{X}, \mu, T)$ is said to be pairwise independently determined (PID) if all pairwise independent d -self joinings ($d \geq 3$) are independent. Each weakly mixing m.p.s. with spectral type singular w.r.t Lebesgue measure is PID (Host 1991); and so is each finite-rank mixing transformation (Ryzhikov 1993).

Using topological model, one can show the following result.

Theorem 4.18 (Gutman et al. 2018) *Let $(X, \mathcal{X}, \mu, T)$ be a weakly mixing and pairwise independently determined (PID) m.p.s. Then for all $d \in \mathbb{N}$ and all $f_1, \dots, f_d \in L^\infty(X, \mathcal{X}, \mu)$, (4) exists a.s. Moreover, for a generic m.p.t. (4) exists a.s.*

Now we discuss the convergence along the cube groups. Host and Kra showed the following convergence along the cube groups.

Theorem 4.19 (Host and Kra 2005) *Let $(X, \mathcal{X}, \mu, T)$ be an m.p.s., and $d \in \mathbb{N}$. Then for functions $f_\varepsilon \in L^\infty(\mu)$, $\varepsilon \in \{0, 1\}^d$, $\varepsilon \neq (0, \dots, 0)$, the averages*

$$\prod_{i=1}^d \frac{1}{N_i - M_i} \cdot \sum_{\mathbf{n} \in [M_1, N_1) \times \dots \times [M_d, N_d)(0, \dots, 0) \neq \varepsilon \in \{0, 1\}^d} \prod f_\varepsilon(T^{\mathbf{n} \cdot \varepsilon} x) \quad (5)$$

converge in $L^2(X)$ as $N_1 - M_1, N_2 - M_2, \dots, N_d - M_d$ tend to $+\infty$.

Let $(X, \mathcal{X}, \mu, T)$ be an ergodic m.p.s. and $d \in \mathbb{N}$. Using Weiss's theorem we can prove that it has a strictly ergodic model $(\hat{X}, \hat{T})$ such that $(\mathbf{Q}^{[d]}(\hat{X}), \mathcal{G}^{[d]})$ is strictly ergodic. Then we can reprove that (5) converge μ a.e. (Huang et al. 2017b). This result for the averages along $[0, N-1]^d$ was first built by Assani (2010), and Chu and Franzikakis (2012) extended the result to a very general case.

It should be noted that these methods were further developed in Donoso and Sun (2016, 2018a, b).

Further Directions

Pointwise Convergence of the Multiple Averages

In contrast to the progress on mean convergence, the problem of pointwise convergence for multiple averages along arithmetic progressions remains open. At present for a single transformation, the most well-known results are due to Bourgain (1990) and Huang et al. (2019d). Donoso and Sun extended the results in Huang et al. (2019d) to commuting transformations in Donoso and Sun (2018a).

Analogue Results

It is always an interesting question to get analogue results in topological dynamics and ergodic theory. We mention two of them:

We say (X, T) is *totally minimal* if (X, T^n) is minimal for and $n \neq 0$. A deep open question is the so-called “odd recurrent problem” (for the ergodic version, see (Frantzikinakis 2004)). Let (X, T) be *totally minimal*. For given $d \in \mathbb{N}$ and $j \in \mathbb{Z}_+$ with $0 \leq j \leq d - 1$, does there exist a sequence $\{n_i\}$ with $n_i \pmod{d} = j$ and $x \in X$ such that

$$T^{n_i}x \rightarrow x, T^{2n_i}x \rightarrow x, \dots, T^{dn_i}x \rightarrow x, i \rightarrow \infty?$$

A related question is that if (X, T) is totally minimal, is it true that there is $x \in X$ such that $\{T^{n^2}x : n \in \mathbb{Z}\}$ is dense in X ? For a partial solution and the ergodic versions, see (Huang et al. 2019c; Furstenberg 1981; Bourgain 1989).

Ramsey Type Theorems

An important question in Ramsey theory is to determine which algebraic equations, or systems of equations, are partition regular over the natural numbers. Partition regularity of the equation $p-(x_1, x_2, \dots, x_n) = 0 (n \geq 2)$ amounts to saying that, for any partition of $\mathbb{N}$ into finitely many cells, some cell contains distinct $x_1, x_2, \dots, x_n$ that

satisfy the equation. The case where the polynomial p is linear was completely solved by Rado (1933). A notorious old question of Erdős and Graham (1980) is that whether the equation $x^2 + y^2 = z^2$ is partition regular?

Hindman-Graham in 1979 asked the following question: for any finite coloring of $\mathbb{N}$, does there exist $x, y \in \mathbb{N}$ such that $\{x, y, xy, x + y\}$ is monochromatic? Moreira (2017) solved a weaker form by showing that: for any finite coloring of $\mathbb{N}$, there exist $x, y \in \mathbb{N}$ such that $\{x, xy, x + y\}$ is monochromatic.

Sarnak Conjecture

The Möbius function $\mu: \mathbb{N} \rightarrow \{-1, 0, 1\}$ is defined by $\mu(1) = 1$ and

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n \text{ is a product of } k \\ & \text{distinct primes;} \\ 0 & \text{otherwise.} \end{cases}$$

Let (X, T) be a t.d.s. We say a sequence ξ is *realized* in (X, T) if there is an $f \in C(X)$ and an $x \in X$ such that $\xi(n) = f(T^n x)$ for any $n \in \mathbb{N}$. A sequence ξ is called *deterministic* if it is realized in a system with zero topological entropy. Here is the conjecture by Sarnak (2009):

The Möbius function μ is linearly disjoint from any deterministic sequence ξ . That is,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(n) \xi(n) = 0.$$

Though there are many papers which solved some special classes of zero entropy systems (see (Green and Tao 2010; Ferenczi et al. 2018; Kułaga-Przymus and Lemańczyk 2019) and the references therein), it seems that there is a long way to go to settle the conjecture completely.

References

- Adler RL, Konheim AG, McAndrew MH (1965) Topological entropy. Trans Am Math Soc 114:309–319
- Akin E (1997) Recurrence in topological dynamics, Furstenberg families and Ellis actions. Plenum Press, New York/London

- Akin E, Kolyada S (2003) Li-Yorke sensitivity. *Non-linearity* 16:1421–1433
- Akin E, Auslander J, Berg K (1993) When is a transitive map chaotic? In: *Convergence in ergodic theory and probability* (Columbus, OH, 1993), 25C40. Ohio State University Mathematical Research Institute Publications, 5. de Gruyter, Berlin, 1996
- Akin E, Glasner E, Huang W, Shao S, Ye X (2010) Sufficient conditions under which a transitive system is chaotic. *Ergodic Theory Dyn Syst* 30:1277–1310
- Assani I (1998) Multiple recurrence and almost sure convergence for weakly mixing dynamical systems. *Israel J Math* 103:111–124
- Assani I (2010) Pointwise convergence of ergodic averages along cubes. *J Anal Math* 110:241–269
- Auslander J (1988) Minimal flows and their extensions. North-Holland mathematics studies, vol 153. North-Holland, Amsterdam
- Auslander J, Furstenberg H (1994) Product recurrence and distal points. *Trans Am Math Soc* 343:221–232
- Auslander J, Yorke J (1980) Interval maps, factors of maps, and chaos. *Thoku Math J* 32(2):177–188
- Bergelson V (2006) Combinatorial and diophantine applications of ergodic theory. Appendix A by A. Leibman and Appendix B by Anthony Quas and Máté Wierdl. In: *Handbook of dynamical systems*, vol 1B. Elsevier B. V., Amsterdam, pp 745–869
- Bergelson V, Leibman A (1996) Polynomial extensions of van der Waerden’s and Szemerédi’s theorems. *J AMS* 9:725–753
- Bergelson V, Leibman A (2002) A nilpotent Roth theorem. *Invent Math* 147:429–470
- Bergelson V, Host B, Kra B (2005) Multiple recurrence and nilsequences. With an appendix by Imre Ruzsa. *Invent Math* 160:261–303
- Billingsley P (1965) *Ergodic theory and information*. Wiley, New York/London/Sydney
- Birkhoff GD (1927) *Dynamical systems*, vol 9. American Mathematical Society Colloquium Publications, New York
- Blanchard F (1992) Fully positive topological entropy and topological mixing. *Symbolic dynamics and its applications*. *AMS Contemp Math* 135:95–105
- Blanchard F (1993) A disjointness theorem involving topological entropy. *Bull Soc Math France* 121:465–478
- Blanchard F, Huang W (2008) Entropy sets, weakly mixing sets and entropy capacity. *Discrete Contin Dyn Syst* 20(2):275–311
- Blanchard F, Lacroix Y (1993) Zero-entropy factors of topological flows. *Proc Am Math Soc* 119:985–992
- Blanchard F, Host B, Maass A, Martinez S, Rudolph DJ (1995) Entropy pairs for a measure. *Ergodic Theory Dyn Syst* 15(4):621–632
- Blanchard F, Glasner E, Host B (1997) A variation on the variational principle and applications to entropy pairs. *Ergodic Theory Dyn Syst* 17:29–43
- Blanchard F, Host B, Maass A (2000) Topological complexity. *Ergodic Theory Dyn Syst* 20:641–662
- Blanchard F, Glasner E, Kolyada S, Maass A (2002a) On Li-Yorke pairs. *J Reine Angew Math* 547:51–68
- Blanchard F, Host B, Ruette S (2002b) Asymptotic pairs in positive-entropy systems. *Ergodic Theory Dyn Syst* 22:671–686
- Bourgain J (1989) Pointwise ergodic theorems for arithmetic sets. With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein. *Inst Hautes Études Sci Publ Math* 69:5–45
- Bourgain J (1990) Double recurrence and almost sure convergence. *J Reine Angew Math* 404:140–161
- Bowen R (1971) Entropy for group endomorphisms and homogeneous spaces. *Trans Am Math Soc* 153:401–414
- Bronstein IU (1979) Extensions of minimal transformation groups. Martinus Nijhoff Publications, The Hague
- Buczolich Z, Mauldin RD (2010) Divergent square averages. *Ann of Math* 171(2):1479–1530
- Cadre B, Jacob P (2005) On pairwise sensitivity. *J Math Anal Appl* 309:375–382
- Cai F, Shao S (2019) Topological characteristic factors along cubes of minimal systems. *Discrete Contin Dyn Syst* 39(9):5301–5317
- Candela P (2017a) Notes on nilspaces: algebraic aspects. *Discrete Anal* 15
- Candela P (2017b) Notes on compact nilspaces. *Discrete Anal* 16
- Cánovas JS (2004) Topological sequence entropy of interval maps. *Nonlinearity* 17(1):49–56
- Chu Q, Frantzikinakis N (2012) Pointwise convergence for cubic and polynomial ergodic averages of non-commuting transformations. *Ergodic Theory Dyn Syst* 32:877–897
- Conze J, Lesigne E (1984) Théorèmes ergodiques pour des mesures diagonales. (French) [Ergodic theorems for diagonal measures]. *Bull Soc Math France* 112:143–175
- Conze J, Lesigne E (1987) Sur un théorème ergodique pour des mesures diagonales. (French) [On an ergodic theorem for diagonal measures]. *Probabilités, 1–31*, Publ. Inst. Rech. Math. Rennes, 1987-1, Université Rennes I, Rennes, 1988
- Conze J, Lesigne E (1988) Sur un théorème ergodique pour des mesures diagonales. (French) [On an ergodic theorem for diagonal measures]. *C R Acad Sci Paris Sr I Math* 306:491–493
- Cornfeld IP, Fomin SV, Sinai YG (1982) *Ergodic theory. Fundamental principles of mathematical sciences*, vol 245. Springer, New York, p x,486
- Coronel A, Maass A, Shao S (2009) Sequence entropy and rigid σ -algebras. *Stud Math* 194:207–230
- de la Rue T (2012) Joinings in ergodic theory. *Mathematics of complexity and dynamical systems*, vol 1–3. Springer, New York, pp 796–809
- de Vries J (1993) *Elements of topological dynamics*. Kluwer, Dordrecht
- Denker M, Grillenberger C, Sigmund K (1976) *Ergodic theory on compact spaces*. Springer lecture notes in mathematics, vol 527. Springer, Berlin

- Derrien J, Lesigne E (1996) Un théorème ergodique polynomial ponctuel pour les endomorphismes exacts et les K-systèmes. (French) [A pointwise polynomial ergodic theorem for exact endomorphisms and K-systems]. *Ann Inst H Poincaré Probab Stat* 32:765–778
- Devaney RL (1989) An introduction to chaotic dynamical systems. Addison-Wesley Publishing Company Advanced Book Program, Redwood City
- Dinaburg EI (1971) A connection between various entropy characterizations of dynamical systems. (Russian) *Izv Akad Nauk SSSR Ser Mat* 35:324–366
- Dong P, Shao S, Ye X (2012) Product recurrent properties, disjointness and weak disjointness. *Israel J Math* 188:463–507
- Dong P, Donoso S, Maass A, Shao S, Ye X (2013) Infinite-step nilsystems, independence and complexity. *Ergodic Theory Dyn Syst* 33:118–143
- Donoso S, Shao S (2017) Uniformly rigid models for rigid actions. *Stud Math* 236:13–31
- Donoso S, Sun W (2016) A pointwise cubic average for two commuting transformations. *Israel J Math* 216:657–678
- Donoso S, Sun W (2018a) Pointwise multiple averages for systems with two commuting transformations. *Ergodic Theory Dyn Syst* 38(6):2132–2157
- Donoso S, Sun W (2018b) Pointwise convergence of some multiple ergodic averages. *Adv Math* 330:946–996
- Dou D, Ye X, Zhang G (2006) Entropy sequences and maximal entropy sets. *Nonlinearity* 19:53–74
- Downarowicz T (2011) Entropy in dynamical systems. *New mathematical monographs*, vol 18. Cambridge University Press, Cambridge
- Downarowicz T (2014) Positive topological entropy implies chaos DC2. *Proc Am Math Soc* 142(1):137–149
- Downarowicz T, Ye X (2002) When every point is either transitive or periodic. *Colloq Math* 93:137–150
- Eberlein E (1975) On topological entropy of semigroups of commuting transformations. In: *International conference on dynamical systems in mathematical physics*, Rennes, 1975, pp 17–62
- Ellis R (1969) *Lectures on topological dynamics*. W. A. Benjamin, Inc., New York
- Ellis R (1973) The Veech structure theorem. *Trans Am Math Soc* 186:203–218
- Ellis R, Gottschalk W (1960) Homomorphisms of transformation groups. *Trans Am Math Soc* 94:258–271
- Ellis R, Keynes H (1971) A characterization of the equicontinuous structure relation. *Trans Am Math Soc* 161:171–181
- Ellis R, Keynes H (1972) Bohr compactifications and a result of Følner. *Israel J Math* 12:314–330
- Ellis R, Nerurkar M (1988) Enveloping semigroup in ergodic theory and a proof of Moore's ergodicity theorem. *Dynamical systems (College Park, MD, 1986–87)*. *Lecture notes in mathematics*, 1342. Springer, Berlin, pp 172–179
- Ellis R, Nerurkar M (1989) Weakly almost periodic flows. *Trans Am Math Soc* 313:103–119
- Ellis R, Glasner S, Shapiro L (1975) Proximal-isometric flows. *Adv Math* 17:213–260
- Erdős P, Graham RL (1980) Old and new problems and results in combinatorial number theory, *Monographies de L'Enseignement Mathématique [Monographs of L'Enseignement Mathématique]*, vol 28. Université de Genève, L'Enseignement Mathématique, Geneva
- Erdős P, Turán P (1936) On some sequences of integers. *J Lond Math Soc* 11:261–264
- Fayad B, Kanigowski A (2016) Multiple mixing for a class of conservative surface flows. *Invent Math* 203(2):555–614
- Ferenczi S, Kulaga-Przymus J, Lemanczyk M (2018) Sarnak's conjecture: what's new. *Lecture Notes Math* 2213:163–235
- Følner E (1954) Generalization of a theorem of Bogoliuboff to topological abelian groups. *Math Scand* 2:5–19
- Frantzikinakis N (2004) The structure of strongly stationary systems. *J Anal Math* 93:359C388
- Frantzikinakis N (2017) Ergodicity of the Liouville system implies the Chowla conjecture. *Discrete analysis paper no. 19*, 41 pp
- Frantzikinakis N, McCutcheon R (2012) Ergodic theory: recurrence. *Mathematics of complexity and dynamical systems*, vol 1–3. Springer, New York, pp 357–368
- Fuhrmann G, Glasner E, Jäger T, Oertel C (2018) Irregular model sets and tame dynamics. *arXiv:1811.06283*
- Furstenberg H (1963) The structure of distal flows. *Am J Math* 85:477–515
- Furstenberg H (1967) Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math Syst Theory* 1:1–49
- Furstenberg H (1977) Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J Anal Math* 31:204–256
- Furstenberg H (1981) Recurrence in ergodic theory and combinatorial number theory. *M. B. Porter lectures*. Princeton University Press, Princeton
- Furstenberg H (1982) IP-systems in ergodic theory. *Conference in modern analysis and probability (New Haven, Conn., 1982)*. *Contemporary Mathematics* 26, American Mathematical Society, Providence, 1984, pp 131–148
- Furstenberg H (1988) Nonconventional ergodic averages. The legacy of John von Neumann (Hempstead, NY, 1988). In: *Proceedings of Symposia in Pure Mathematics*, vol 50, American Mathematical Society, Providence, 1990, pp 43–56
- Furstenberg H (1991) Recurrent ergodic structures and Ramsey theory. In: *Proceedings of the International Congress of Mathematicians*, vol I, II (Kyoto, 1990). *Mathematical Society of Japan*, Tokyo, 1991, pp 1057–1069
- Furstenberg H (2010) Ergodic structures and non-conventional ergodic theorems. *International Congress of Mathematicians*
- Furstenberg H, Weiss B (1977) The finite multipliers of infinite ergodic transformations. The structure of attractors in dynamical systems. In: *Proceedings of the conference on North Dakota State University*, Fargo, N.D., 1977. *Lecture notes in mathematics*, 668. Springer, Berlin, 1978, pp 127–132

- Furstenberg H, Weiss B (1978) Topological dynamics and combinatorial number theory. *J Anal Math* 34:61–85
- Furstenberg H, Weiss B (1996) A mean ergodic theorem for $\frac{1}{N} \sum_{n=1}^N f(T^n x)g(T^{m^2} x)$. In: *Convergence in ergodic theory and probability* (Columbus, OH, 1993). Ohio State University Mathematical Research Institute Publications, S. de Gruyter, Berlin, pp 193–227
- García-Ramos F (2017) Weak forms of topological and measure theoretical equicontinuity: relationships with discrete spectrum and sequence entropy. *Ergodic Theory Dyn Syst* 37:1211–1237
- García-Ramos F, Jäger T, Ye X (2019) Mean equicontinuity, almost automorphy and regularity, preprint
- Glasner S (1976) Proximal flows. *Lecture notes in mathematics*, vol 517. Springer, Berlin/New York
- Glasner E (1994) Topological ergodic decompositions and applications to products of powers of a minimal transformation. *J Anal Math* 64:241–262
- Glasner E (1996) Structure theory as a tool in topological dynamics. In: *Descriptive set theory and dynamical systems* (Marseille-Luminy, 1996), 173C209. London Mathematical Society lecture note series, vol 277. Cambridge University Press, Cambridge, 2000
- Glasner E (1997) A simple characterization of the set of μ -entropy pairs and applications. *Israel J Math* 102:13–27
- Glasner E (1998) On minimal actions of Polish groups. *Topology Appl* 117:259–272
- Glasner E (2003) *Ergodic theory via Joinings*. Mathematical surveys and monographs, 101. American Mathematical Society, Providence
- Glasner E (2018) The structure of tame minimal dynamical systems for general groups. *Invent Math* 211(1):213–244
- Glasner S, Maon D (1989) Rigidity in topological dynamics. *Ergodic Theory Dyn Syst* 9:309–320
- Glasner S, Weiss B (1983) Minimal transformations with no common factor need not be disjoint. *Israel J Math* 45:1–8
- Glasner E, Weiss B (1994) Strictly ergodic, uniform positive entropy models. *Bull Soc Math France* 122(3):399–412
- Glasner E, Weiss B (1995a) Quasi-factors of zero-entropy systems. *J Am Math Soc* 8(3):665–686
- Glasner E, Weiss B (1995b) Topological entropy of extensions. In: *Ergodic theory and its connections with harmonic analysis* (Alexandria, 1993). London Mathematical Society lecture note series, vol 205. Cambridge University Press, Cambridge, pp 299–307
- Glasner E, Weiss B (2005) On the interplay between measurable and topological dynamics. In: Hasselblatt B, Katok A (eds) *Handbook of dynamical systems*, vol 1B. North-Holland, Amsterdam, pp 597–648
- Glasner E, Weiss B (2015) On doubly minimal systems and a question regarding product recurrence. [arXiv:1508.02817](https://arxiv.org/abs/1508.02817)
- Glasner E, Ye X (2009) Local entropy theory. *Ergodic Theory Dyn Syst* 29:321–356
- Glasner E, Gutman Y, Ye X (2018) Higher order regionally proximal equivalence relations for general minimal group actions. *Adv Math* 333:1004–1041
- Glasner E, Huang W, Shao S, Ye X (2019) Regionally proximal relation of order d along arithmetic progressions and nilsystems, in prepare
- Goodman TNT (1971) Relating topological entropy and measure entropy. *Bull Lond Math Soc* 3:176–180
- Goodman TNT (1974) Topological sequence entropy. *Proc Lond Math Soc* 29:331–350
- Goodwyn L (1969) Topological entropy bounds measure-theoretic entropy. *Proc Am Math Soc* 23:679–688
- Gottschalk W (n.d.) Orbit closure decompositions and almost periodic properties. *Bull Am Math Soc* 50:915–919
- Gottschalk W, Hedlund G (1955) *Topological dynamics*. American Mathematical Society Colloquium publications, vol 36. American Mathematical Society, Providence
- Green B, Tao T (2008) The primes contain arbitrarily long arithmetic progressions. *Ann Math* 167(2):481–547
- Green B, Tao T (2010) The Möbius function is strongly orthogonal to nilsequences. *Ann Math* 175(2):541–566
- Gutman Y, Huang Y, Shao S, Ye X (2018) Almost sure convergence of the multiple ergodic average for certain weakly mixing systems. *Acta Math Sinica* 34:79–90
- Gutman Y, Manners F, Varjú P (2019a) The structure theory of Nilspaces III: inverse limit representations and topological dynamics. [http://arxiv.org/abs/1605.08950](https://arxiv.org/abs/1605.08950)
- Gutman Y, Manners F, Varjú P (2019b) The structure theory of Nilspaces II: representation as nilmanifolds. *Trans Am Math Soc* 371:4951–4992
- Gutman Y, Manners F, Varjú P (2020) The structure theory of Nilspaces I. *J Anal Math* 140:299–369
- Haddad K, Ott W (2008) Recurrence in pairs. *Ergodic Theory Dyn Syst* 28:1135–1143
- Halmos PR (1960) *Lectures on Ergodic theory*. Chelsea Publishing Co, New York, p vii, 101
- Halmos P, von Neumann J (1942) Operator methods in classical mechanics II. *Ann Math* 43:332–350
- Hindman N (1974) Finite sums from sequences within cells of a partition of $\mathbb{N}$. *J Combin Theory Ser A* 17:1–11
- Hochman M (2012) On notions of determinism in topological dynamics. *Ergodic Theory Dyn Syst* 32:119–140
- Host B (1991) Mixing of all orders and pairwise independent joinings of systems with singular spectrum. *Israel J Math* 76:289–298
- Host B, Kra B (2001) Convergence of Conze-Lesigne averages. *Ergodic Theory Dyn Syst* 21:493–509
- Host B, Kra B (2002) An odd Furstenberg-Szemerédi theorem and quasi-affine systems. *J Anal Math* 86:183–220
- Host B, Kra B (2005) Nonconventional averages and nilmanifolds. *Ann Math* 161:398–488
- Host B, Kra B (2008) Parallelepipedes, nilpotent groups and Gowers norms. *Bull Soc Math France* 136(3):405–437
- Host B, Kra B (2018) *Nilpotent structures in Ergodic theory*. Mathematical surveys and monographs, 236. American Mathematical Society, Providence

- Host B, Maass A (2007) Nilsystèmes d'ordre deux et parallélépipèdes. *Bull Soc Math France* 135:367–405
- Host B, Kra B, Maass A (2010) Nilsequences and a structure theory for topological dynamical systems. *Adv Math* 224:103–129
- Host B, Kra B, Maass A (2016) Variations on topological recurrence. *Monatsh Math* 179:57–89
- Huang W (2006) Tame systems and scrambled pairs under an abelian group action. *Ergodic Theory Dyn Syst* 26(5):1549–1567
- Huang W (2008) Stable sets and ε -stable sets in positive-entropy systems. *Commun Math Phys* 279(2):535–557
- Huang W, Jin L (2016) Stable sets and mean Li-Yorke chaos in positive entropy actions of bi-orderable amenable groups. *Ergodic Theory Dyn Syst* 36(8):2482–2497
- Huang W, Lu K (2017) Entropy, chaos, and weak horseshoe for infinite-dimensional random dynamical systems. *Commun Pure Appl Math* 70(10):1987–2036
- Huang W, Ye X (2002a) Devaney's chaos or 2-scattering implies Li-Yorke's chaos. *Topology Appl* 117:259–272
- Huang W, Ye X (2002b) An explicit scattering, non-weakly mixing example and weak disjointness. *Nonlinearity* 15:1–14
- Huang W, Ye X (2004) Topological complexity, return times and weak disjointness. *Ergodic Theory Dyn Syst* 24(3):825–846
- Huang W, Ye X (2005) Dynamical systems disjoint from any minimal system. *Trans Am Math Soc* 357(2):669–694
- Huang W, Ye X (2006) A local variational relation and applications. *Israel J Math* 151:237–280
- Huang W, Ye X (2009) Combinatorial lemmas and applications to dynamics. *Adv Math* 220(6):1689–1716
- Huang W, Ye X (2012) Generic eigenvalues, generic homomorphisms and weak disjointness. *Contemp Math* 567:119–143
- Huang W, Li S, Shao S, Ye X (2003) Null systems and sequence entropy pairs. *Ergodic Theory Dyn Syst* 23(5):1505–1523
- Huang W, Maass A, Ye X (2004) Sequence entropy pairs and complexity pairs for a measure. *Ann Inst Fourier (Grenoble)* 54(4):1005–1028
- Huang W, Shao S, Ye X (2005) Mixing via sequence entropy. Algebraic and topological dynamics. *Contemp Math* 385:101–122
- Huang W, Ye X, Zhang G (2006) A local variational principle for conditional entropy. *Ergodic Theory Dyn Syst* 26(1):219–245
- Huang W, Ye X, Zhang G (2007) Relative entropy tuples, relative U.P.E. and C.P.E. extensions. *Israel J Math* 158:249–283
- Huang W, Lu P, Ye X (2011a) Measure-theoretical sensitivity and equicontinuity. *Israel J Math* 183:233–283
- Huang W, Ye X, Zhang G (2011b) Local entropy theory for a countable discrete amenable group action. *J Funct Anal* 261(4):1028–1082
- Huang W, Li J, Ye X (2014) Stable sets and mean Li-Yorke chaos in positive entropy systems. *J Funct Anal* 266(6):3377–3394
- Huang W, Xu L, Yi YF (2015) Asymptotic pairs, stable sets and chaos in positive entropy systems. *J Funct Anal* 268(4):824–846
- Huang W, Shao S, Ye X (2016) Nil Bohr-sets and almost automorphy of higher order. *Memoirs of the American Mathematical Society*. American Mathematical Society, Providence
- Huang W, Li J, Ye X, Zhou XY (2017a) Positive topological entropy and Δ -weakly mixing sets. *Adv Math* 306:653–683
- Huang W, Shao S, Ye X (2017b) Strictly ergodic models and pointwise ergodic averages for cubes. *Comm Math Stat* 5:93–122
- Huang W, Li J, Thouvenot J, Xu L, Ye X (2019a) Bounded complexity, mean equicontinuity and discrete spectrum. *Ergodic Theory Dyn Syst*. arXiv:1806.02980
- Huang W, Lian Z, Shao S, Ye X (2019b) Minimal systems with finitely many ergodic measures, preprint
- Huang W, Shao S, Ye X (2019c) Topological correspondence of multiple ergodic averages of nilpotent group actions. *J Anal Math*. arXiv:1604.07113
- Huang W, Shao S, Ye X (2019d) Pointwise convergence of multiple ergodic averages and strictly ergodic models. *J Anal Math*. arXiv:1406.5930
- Huang W, Shao S, Ye X (2019e) An answer to Furstenberg's problem on topological disjointness. *Ergodic Theory Dyn Syst*. arXiv:1807.10155
- Huang W, Wang Z, Ye X (2019f) Measure complexity and Möbius disjointness. *Adv Math* 347:827–858
- Hulse P (1982) Sequence entropy and subsequence generators. *J Lond Math Soc* 26:441–450
- Hulse P (1986) Sequence entropy relative to an invariant σ -algebra. *J Lond Math Soc* 33:59–72
- James J, Koberda T, Lindsey K, Silva CE, Speh P (2008) Measurable sensitivity. *Proc Am Math Soc* 136:3549–3559
- Jewett RI (1969/1970) The prevalence of uniquely ergodic systems. *J Math Mech* 19:717–729
- Kalikow S (1984) Twofold mixing implies threefold mixing for rank one transformations. *Ergodic Theory Dyn Syst* 4:237–259
- Kanigowski A, Kułaga-Przymus J, Ułcigrai C (2017) Multiple mixing and parabolic divergence in smooth area-preserving flows on higher genus surfaces. *J Eur Math Soc*. arXiv: 1606.09189
- Katok A (1980) Lyapunov exponents, entropy and the periodic orbits for diffeomorphisms. *Inst Hautes études Sci Publ Math* 51:137–173
- Katok A, Hasselblatt B (1995) Introduction to the modern theory of dynamical systems. In: *Encyclopedia of mathematics and its applications*, vol 54. Cambridge University Press, Cambridge, p xviii,802
- Katznelson Y (2001) Chromatic numbers of Cayley graphs on $\mathbb{Z}$ and recurrence. *Combinatorica* 21:211–219
- Kerr D, Li H (2007) Independence in topological and C^* -dynamics. *Math Ann* 338(4):869–926
- Kerr D, Li H (2009) Combinatorial independence in measurable dynamics. *J Funct Anal* 256(5):1341–1386
- Kolmogorov AN (1958) A new metric invariant of transient dynamical systems and automorphisms in

- Lebesgue spaces. (Russian) Dokl Akad Nauk SSSR (NS) 119:861–864
- Kolmogorov AN (1959) Entropy per unit time as a metric invariant of automorphisms. (Russian) Dokl Akad Nauk SSSR 124:754–755
- Kolyada S, Snoha L (1997) Some aspects of topological transitivity I a survey. In: Iteration theory (ECIT 94) (Opava), Grazer mathematische Berichte, vol 334. Karl-Franzens-Universität Graz, Graz, p 3C35
- Korner TW (1987) Recurrence without uniform recurrence. Ergodic Theory Dyn Syst 7:559–566
- Kra B (2006) From combinatorics to ergodic theory and back again. International Congress of Mathematicians, vol III. European Mathematical Society, Zürich, pp 57–76
- Krieger W (1970/1971) On unique ergodicity. In: Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, vol II: probability theory. University of California Press, Berkeley, 1972, pp 327–346
- Kronecker L. (n.d.) Naherrungsweise ganzzahlige Auflosunglinear Gleichungen, S.-B. Preuss, Akad. Wiss., 1179–93, 1271–99. Werke III (1), pp 47–109
- Krug E, Newton D (1972) On sequence entropy of automorphisms of a Lebesgue space. Z Wahrscheinlichkeitstheorie Verw Gebiete 24:211–214
- Kuřaga-Przymus J, Lemańczyk M (2019) Sarnak’s conjecture from the ergodic theory point of view. This volume
- Kurka P (2003) Topological and symbolic dynamics. Cours Spécialisés specialized courses, 11. Société Mathématique de France, Paris
- Kushnirenko AG (1967) On metric invariants of entropy type. Russ Math Surv 22:53–61
- Kvietniak D, Oprocha P (2012) On weak mixing, minimality and weak disjointness of all iterates. Ergodic Theory Dyn Syst 32:1661–1672
- Lehrer E (1987) Topological mixing and uniquely ergodic systems. Israel J Math 57:239–255
- Leibman A (2005) Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold. Ergodic Theory Dyn Syst 25:201–213
- Lesigne E (1989) Théorèmes ergodiques pour une translation sur un nilvariété. (French) [Ergodic theorems for a translation on a nilmanifold]. Ergodic Theory Dyn Syst 9:115–126
- Li J, Qiao YX (2018) Mean Li-Yorke chaos along some good sequences. Monatsh Math 186(1):153–173
- Li J, Ye X (2016) Recent development of chaos theory in topological dynamics. Acta Math Sinica 32:83–114
- Li TY, Yorke J (1975) Period three implies chaos. Am Math Mon 82:985–992
- Li J, Tu S, Ye X (2015a) Mean equicontinuity and mean sensitivity. Ergodic Theory Dyn Syst 35:2587–2612
- Li J, Yan K, Ye X (2015b) Recurrence properties and disjointness on the induced spaces. Discrete Contin Dyn Syst 35:1059–1073
- Lindenstrauss E (1995) Lowering topological entropy. J Anal Math 67:231–267
- Lindenstrauss E (1999) Measurable distal and topological distal systems. Ergodic Theory Dyn Syst 19:1063–1076
- Liu K (2019) Δ -weakly mixing subset in positive entropy actions of a nilpotent group. J Differ Equ 267:525–546
- Maass A, Shao S (2007) Structure of bounded topological-sequence-entropy minimal systems. J Lond Math Soc 76(3):702–718
- McMahon DC (1976) Weak mixing and a note on a structure theorem for minimal transformation groups. Ill J Math 20:186–197
- McMahon DC (1978) Relativized weak disjointness and relatively invariant measures. Trans Am Math Soc 236:225–237
- Misiurewicz M (1976) A short proof of the variational principle for a $\mathbb{Z}_+^n$ action on a compact space. Astérisque 40:227–262
- Moothathu TKS (2010) Diagonal points having dense orbit. Colloq Math 120:127–138
- Moreira J (2017) Monochromatic sums and products in $\mathbb{N}$. Ann Math 185:1069–1090
- Nemytskiĭ VV (1949) Topological problems in the theory of dynamical systems, Uspekhi Matematicheskikh Nauk, 5. American Mathematical Society translation, no. 103, 1954
- Nemytskiĭ VV, Stepanov VV (1949) Qualitative theory of differential equations. GITTL Moscow (Russian). Princeton University Press, Princeton, 1960
- Nicol M, Petersen K (2012) Ergodic theory: basic examples and constructions. Mathematics of complexity and dynamical systems, vol 1–3. Springer, New York, pp 264–287
- Oprocha P (2010) Weak mixing and product recurrence. Ann Inst Fourier (Grenoble) 60(4):1233–1257
- Oprocha P (2019) Double minimality, entropy and disjointness with all minimal systems. Discrete Contin Dyn Syst 39:263–275
- Oprocha P, Zhang G (2013) On weak product recurrence and synchronization of return times. Adv Math 244:395–412
- Oprocha P, Zhang G (2014) Topological aspects of dynamics of pairs, tuples and sets. In: Recent progress in general topology. III. Atlantis Press, Paris, pp 665–709
- Ornstein D (1970) Bernoulli shifts with the same entropy are isomorphic. Adv Math 4:337–352
- Park KK, Siemaszko A (2001) Relative topological Pinsker factors and entropy pairs. Monatsh Math 134:67–79
- Parry W (1967) Zero entropy of distal and related transformations. In: Auslander J, Gottschalk W (eds) Topological dynamics. Benjamin, New York
- Pinsker MS (1960) Dynamical systems with completely positive or zero entropy. Dokl. Akad. Nauk SSSR 133:1025–1026 (Russian); translated as Soviet Math. Dokl. 1:937–938
- Qiu J, Zhao J (2020) A note on mean equicontinuity. J Dyn Diff Equat 32:101–116
- Qiu J, Zhao J. (n.d.) Null systems in the non-minimal case. Ergodic Theory Dyn Syst. <https://www.cambridge.org/core/journals/ergodic-theory-and-dynamical-systems/article/null-systems-in-the-nonminimal-case/F8AD1510DBAE43CA94CC0CB5A3DD22DD>
- Rado R (1933) Studien zur Kombinatorik (German). Math Z 36:424C470
- Rohlin, VA (1967) Lectures on the entropy theory of transformations with invariant measure. (Russian) Uspekhi Mat. Nauk 22. 137(5):3–56

- Rohlin VA, Sinai YG (1961) Construction and properties of invariant measurable partitions. Dokl Akad Nauk SSSR 141:1038–1041. [In Russian]
- Romagnoli PP (2003) A local variational principle for the topological entropy. Ergodic Theory Dyn Syst 23:1601–1610
- Rudolph D (1979) An example of a measure preserving map with minimal self-joinings, and applications. J Anal Math 35:97–122
- Rudolph DJ (1990) Fundamentals of measurable dynamics. Oxford Science publications. The Clarendon Press Oxford University Press, New York, p x,168
- Ryzhikov VV (1993) Joinings and multiple mixing of the actions of finite rank. (Russian) Funktsional. Anal i Prilozhen 27: 63–78, 96; translation in Funct. Anal Appl 27:128–140
- Saleski A (1977) Sequence entropy and mixing. J Math Anal Appl 60:58–66
- Sarnak P (2009) Three lectures on the Mobius function randomness and dynamics. <http://publications.ias.edu/sarnak/paper/512>
- Schweizer B, Smítal J (1994) Measures of chaos and a spectral decomposition of dynamical systems on the interval. Trans Am Math Soc 344:737–754
- Serafin J (2013) Non-existence of a universal zero-entropy system. Israel J Math 194:349–358
- Shao S, Ye X (2012) Regionally proximal relation of order d is an equivalence one for minimal systems and a combinatorial consequence. Adv Math 231:1786–1817
- Sinai YJ (1959) On the concept of entropy for a dynamic system. (Russian) Dokl Akad Nauk SSSR 124:768–771
- Snoha L, Ye X, Zhang R (2018) Topology and topological sequence entropy. Sci China Math. arXiv:1810.00497
- Szegedy B (2012) On higher order Fourier analysis. arXiv: 1203.2260
- Szemerédi E (1975) On sets of integers containing no k elements in arithmetic progression. Acta Arith 27:199–245
- Tan F (2011) The set of sequence entropies for graph maps. Topology Appl 158(3):533–541
- Tan F, Ye X, Zhang R (2010) The set of sequence entropies for a given space. Nonlinearity 23(1):159–178
- Tao T (2008) Norm convergence of multiple ergodic averages for commuting transformations. Ergodic Theory Dyn Syst 28:657–688
- van der Waerden BL (1927) Beweis einer Baudetschen Vermutung. Nieuw Arch Wisk 15:212–216
- Veech WA (1968) The equicontinuous structure relation for minimal Abelian transformation groups. Am J Math 90:723–732
- Veech WA (1970) Point-distal flows. Am J Math 92:205–242
- Veech WA (1977) Topological systems. Bull Am Math Soc 83:775–830
- Walsh M (2012) Norm convergence of nilpotent ergodic averages. Ann Math 175:1667–1688
- Walters P (1982) An introduction to ergodic theory. Graduate texts in mathematics, 79. Springer, New York
- Wang YP, Chen EC, Zhou XY (2019) Mean Li-Yorke chaos for random dynamical systems. J Diff Equat 267(4):2239–2260
- Ward T (2012) Ergodic theory: interactions with combinatorics and number theory. Mathematics of complexity and dynamical systems, vol 1–3. Springer, New York, pp 313–326
- Weiss B (1985) Strictly ergodic models for dynamical systems. Bull Amer Math Soc (NS) 13:143–146
- Weiss B (1989) Countable generators in dynamics – universal minimal models. Contemp Math 94:321–326
- Weiss B (1995) Multiple recurrence and doubly minimal systems. Topological dynamics and applications (Minneapolis, MN, 1995). Contemporary mathematics, 215. American Mathematical Society, Providence, 199, pp 189–196
- Weiss B (2000) Single orbit dynamics. CBMS regional conference series in mathematics, 95. American Mathematical Society, Providence
- Ye X, Zhang G (2007) Entropy points and applications. Trans Am Math Soc 359:6167–6186
- Ye X, Zhang R (2008) Countable compacta admitting homeomorphisms with positive sequence entropy. J Dyn Diff Equat 20(4):867–882
- Yu T (2019) Measure-theoretic mean equicontinuity and bounded complexity. J Diff Equat 267:6152–6170
- Zhang Q (1992) Sequence entropy and mild mixing. Can J Math 44:215–224
- Zhang Q (1993) Conditional sequence entropy and mild mixing extensions. Can J Math 45:429–448
- Zhang G (2006) Relative entropy, asymptotic pairs and chaos. J Lond Math Soc 73(1):157–172
- Zhou XM, Chen ER, Zhou XY (2017) Relative entropy and mean Li-Yorke chaos. Adv Math (China) 46:407–414
- Ziegler T (2005) A non-conventional ergodic theorem for a nilsystem. Ergodic Theory Dyn Syst 25:1357–1370
- Ziegler T (2007) Universal characteristic factors and Furstenberg averages. J Am Math Soc 20:53–97
- Zimmer RJ (1976a) Extensions of ergodic group actions. Ill J Math 20:373–409
- Zimmer RJ (1976b) Ergodic actions with generalized discrete spectrum. Ill J Math 20:555–588

Symbolic Dynamics

Brian Marcus

Department of Mathematics, University of British Columbia, Vancouver, BC, Canada

Article Outline

Glossary

Definition of the Subject

Introduction

Origins of Symbolic Dynamics: Modeling of Dynamical Systems

Shift Spaces and Sliding Block Codes

Shifts of Finite Type and Sofic Shifts

Entropy and Periodic Points

The Conjugacy Problem

Other Coding Problems

Coding for Data Recording Channels

Connections with Information Theory and Ergodic Theory

Higher Dimensional Shift Spaces

Future Directions

Addendum to the Second Edition

Bibliography

Glossary

Almost conjugacy (Section “[Other Coding Problems](#)”) A common extension of two shift spaces given by factor codes that are one-to-one almost everywhere.

Automorphism (Section “[The Conjugacy Problem](#)”) An invertible sliding block code from a shift space to itself; equivalently, a shift-commuting homeomorphism from a shift space to itself; equivalently, a topological conjugacy from a shift space to itself.

Dimension group (Section “[The Conjugacy Problem](#)”) A particular group associated to a shift of finite type. This group, together with a distinguished sub-semigroup and an

automorphism, captures many invariants of topological conjugacy for shifts of finite type.

Embedding (Section “[Shift Spaces and Sliding Block Codes](#)”) A one-to-one sliding block code from one shift space to another; equivalently, a one-to-one continuous shift-commuting mapping from one shift space to another.

Factor map (Section “[Shift Spaces and Sliding Block Codes](#)”) An onto sliding block code from one shift space to another; equivalently, an onto continuous shift-commuting mapping from one shift space to another. Sometimes called *Factor Code*.

Finite equivalence (Section “[Other Coding Problems](#)”) A common extension of two shift spaces given by finite-to-one factor codes.

Full shift (Section “[Shift Spaces and Sliding Block Codes](#)”) The set of all bi-infinite sequences over an alphabet (together with the shift mapping). Typically, the alphabet is finite.

Higher dimensional shift space (Section “[Higher Dimensional Shift Spaces](#)”) A set of bi-infinite arrays of a given dimension, determined by a collection of finite forbidden arrays. Typically, the alphabet is finite.

Markov partition (Section “[Origins of Symbolic Dynamics: Modeling of Dynamical Systems](#)”) A finite cover of the underlying phase space of a dynamical system, which allows the system to be modeled by a shift of finite type. The elements of the cover are closed sets, which are allowed to intersect only on their boundaries.

Measure of maximal entropy (Section “[Connections with Information Theory and Ergodic Theory](#)”) A shift-invariant measure of maximal measure-theoretic entropy on a shift space. Its measure-theoretic entropy coincides with the topological entropy of the shift space.

Road problem (Section “[Other Coding Problems](#)”) A recently-solved classical problem in symbolic dynamics, graph theory and automata theory.

Run-length limited shift (Section “[Coding for Data Recording Channels](#)”) The set of all bi-infinite binary sequences whose runs of zeros, between two successive ones, are bounded below and above by specific numbers.

Shift equivalence (Section “[The Conjugacy Problem](#)”) An equivalence relation on defining matrices for shifts of finite type. This relation characterizes the corresponding shifts of finite type, up to an eventual notion of topological conjugacy.

Shift space (Section “[Shift Spaces and Sliding Block Codes](#)”) A set of bi-infinite sequences determined by a collection of finite forbidden words; equivalently, a closed shift-invariant subset of a full shift.

Shift of finite type (Section “[Shifts of Finite Type and Sofic Shifts](#)”) A set of bi-infinite sequences determined by a *finite* collection of finite forbidden words.

Sliding block code (Section “[Shift Spaces and Sliding Block Codes](#)”) A mapping from one shift space to another determined by a finite sliding block window; equivalently, a continuous shift-commuting mapping from one shift space to another.

Sofic shift (Section “[Shifts of Finite Type and Sofic Shifts](#)”) A shift space which is a factor of a shift of finite type; equivalently, a set of bi-infinite sequences determined by a finite directed labeled graph.

State splitting (Section “[The Conjugacy Problem](#)”) A splitting of states in a finite directed graph that creates a new graph, whose vertices are the split states. The operation that creates the new graph from the original graph is a basic building block for all topological conjugacies between shifts of finite type.

Strong shift equivalence (Section “[The Conjugacy Problem](#)”) An equivalence relation on defining matrices for shifts of finite type. In principle, this relation characterizes the corresponding shifts of finite type, up to topological conjugacy.

Topological conjugacy (Section “[Shift Spaces and Sliding Block Codes](#)”) A bijective sliding block code from one shift space to another;

equivalently, a shift-commuting homeomorphism from one shift space to another. Sometimes called *conjugacy*.

Topological entropy (Section “[Entropy and Periodic Points](#)”) The asymptotic growth rate of the number of finite sequences of given length in a shift space (as the length goes to infinity).

Zeta function (Section “[Entropy and Periodic Points](#)”) An expression for the number of periodic points of each given period in a shift space.

Definition of the Subject

Symbolic dynamics is the study of shift spaces, which consist of infinite or bi-infinite sequences defined by a shift-invariant constraint on the finite-length subwords. Mappings between two such spaces can be regarded as codes or encodings. Shift spaces are classified, up to various kinds of invertible encodings, by combinatorial, algebraic, topological, and measure-theoretic invariants.

The subject is intimately related to many other areas of research, including dynamical systems, ergodic theory, automata theory and information theory. Shift spaces and their associated shift mappings are used to model a rich and important class of smooth dynamical systems and ergodic measure-preserving transformations. These models have provided a valuable tool for classifying and understanding fundamental properties of dynamical systems. In addition, techniques from symbolic dynamics have had profound applications for data recording applications, such as algorithms and analysis of invertible encodings, and problems in matrix theory, such as characterization of the set of eigenvalues of a nonnegative matrix.

Introduction

This article is intended to give a picture of major topics in symbolic dynamics. Section “[Origins of Symbolic Dynamics: Modeling of Dynamical](#)

Systems” reviews the roots of symbolic dynamics in modeling of dynamical systems. Section “**Shift Spaces and Sliding Block Codes**” lays the foundation by defining the kinds of spaces and mappings considered in the subject. Section “**Shifts of Finite Type and Sofic Shifts**” focuses on distinguished special classes of spaces, known as shifts of finite type and sofic shifts. Section “**Entropy and Periodic Points**” introduces the most fundamental invariants, periodic points and topological entropy. Sections “**The Conjugacy Problem**” and “**Other Coding Problems**” survey progress on the conjugacy problem and other classification/coding problems for shifts of finite type and sofic shifts. In section “**Coding for Data Recording Channels**,” we present applications to coding for data recording. Section “**Connections with Information Theory and Ergodic Theory**” provides a link with information theory and ergodic theory. Finally, section “**Higher Dimensional Shift Spaces**” treats higher dimensional symbolic dynamics.

While this article covers many of the most important topics in the subject, others have been omitted or treated lightly, due to space limitations. These include one-sided shift spaces, countable state symbolic systems, orbit equivalence, flow equivalence, the automorphism group, cellular automata, and substitution systems. References to work in these subareas can be found in the sources mentioned below.

For introductory reading on symbolic dynamics and its applications, beyond this article, one can consult the textbooks Kitchens (1998) and Lind and Marcus (1995). There are also excellent introductory survey articles, such as Boyle (1993), Lind and Schmidt (2002), and S. Williams (2004a). In addition, there are very good expositions which focus on other aspects of the subject. These include Beal (1993), which focuses on connections between symbolic dynamics and automata theory, the lecture notes Marcus-Roth-Siegel (1998), which focusses on constrained coding applications, and Immink (2004a), which focuses on applications to data storage. There are also several excellent collections of articles on special areas of the subject, such as Blanchard et al. (2000), Walters (1992), Williams (2004b).

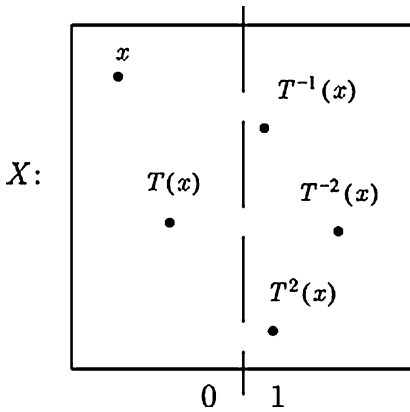
The book (1985) by Berstel and Perrin treats the subject of variable length codes and contains many ideas related to symbolic dynamics. Finally, there is an excellent recent survey by Boyle on Open Problems in Symbolic Dynamics (Boyle 2007).

Origins of Symbolic Dynamics: Modeling of Dynamical Systems

Symbolic dynamics began as an effort to model dynamical systems using sequences of symbols. A *dynamical system* is a pair (X, T) where X is a set and T is a transformation from X to itself. For definiteness, we assume that T is invertible, although this is not a necessary restriction. Since T maps X to itself, we can iterate $T: T^2 = T \circ T$, $T^3 = T \circ T \circ T$, etc. The *orbit* of a point $x \in X$ is the sequence of points: $\dots, T^{-2}(x), T^{-1}(x), x, T(x), T^2(x), \dots$. In the theory of dynamical systems, one asks questions about orbits such as the following: Are there periodic orbits (i.e., x such that $T^n(x) = x$ for some $n > 0$)? Are there dense orbits (the orbit of x is dense if for any point y in X , $T^n(x)$ is “close” to y for some n)? How does the behavior of an orbit vary with x ? How can we describe the collection of all orbits of the dynamical system? When is the dynamical system “chaotic”? For more information on dynamical systems, we refer the reader to (Blanchard et al. 2004; Devaney 1987; Hassellblatt and Katok 1995).

The subject of dynamical systems has its roots in Classical Mechanics; in that setting, X is the set of all possible states of a system (e.g., the positions, momenta of all particles in a physical system), and the transformation T is the time evolution map, which maps the state of the system at one time to the state of the system at one time unit later.

Symbolic dynamics provides a model for the orbits of a dynamical system (X, T) via a space of sequences. This is done by “quantizing” X into cells, associating symbols to the cells and representing points as bi-infinite sequences of symbols. For instance, in Fig. 1, X is a square, and T is some transformation of the square.



Symbolic Dynamics, Fig. 1 Representing points symbolically

We have drawn a portion of the orbit of a point $x \in X$ for the dynamical system (X, T) . We have also quantized X into two cells: the left half, called “0,” and the right half, called “1.” Then the point x is represented by the bi-infinite sequence $s(x) = \dots s_{-2}s_{-1}.s_0s_1s_2\dots$ where s_n is the label of the cell to which $T^n(x)$ belongs (here, we use the decimal point to separate coordinates $s_i, i < 0$ from $s_i, i \geq 0$). So, for x as given in Fig. 1, we see that

$$s(x) = \dots 11.001\dots$$

for instance, $s_0 = 0$ because x belongs to the left half of the square, and $s_2 = 1$ because $T^2(x)$ belongs to the right half of the square. Now, if x is represented by the sequence $s(x)$, then $T(x)$ is represented by the shift of $s(x)$:

$$s(T(x)) = \dots 110.01\dots$$

So, T is represented “symbolically” as the shift transformation.

By representing all points of X as bi-infinite sequences, we obtain a *symbolic dynamical system* (Y, σ) where Y is a set of sequences (representing X) and σ is the shift transformation (representing T). The “symbolic” refers to the symbols, and the “dynamical” refers to the action of the shift transformation.

For this representation to be faithful, distinct points should be represented by distinct sequences, and this imposes extra conditions on

how X is quantized into cells. Also, Y is typically a set of sequences constrained by certain rules, such as a certain symbol may only be followed by certain other symbols.

In this way, one can use symbolic dynamics to study dynamical systems. Properties of orbits of the original dynamical system are reflected in properties of the resulting sequences. For instance, a point whose orbit is periodic becomes a periodic sequence, and the distribution of the orbit of a point x in X is reflected in the distribution of finite strings within $s(x)$. Beginning with Hadamard (1898) in 1898 and followed by Hedlund, Morse and others in the 1920s, 1930s, and 1940s (Hedlund 1939, 1944; Morse and Hedlund 1938, 1940), this method was used to prove the existence of periodic, almost periodic and other interesting motions in classical dynamical systems, such as geodesic flows on surfaces of negative curvature; this was done by finding interesting sequences satisfying the constraints defined by the corresponding symbolic dynamical system. Later on, this was extended to general hyperbolic systems, where the symbolic dynamics is constructed using a *Markov Partition*, which is a disjoint collection of open sets whose closures cover X , each of which looks like a “rectangle”, with vertical (resp., horizontal) fibers contracted (resp., expanded) by T . Markov Partitions were developed by Adler and Weiss (1970), Sinai (1968), and Bowen (1970, 1973); see also Bedford (1986) and Sect. 6.5 in Lind and Marcus (1995).

In more recent years, symbolic dynamics has been used as a tool in classification problems for dynamical systems. Here, the problem of determining when one dynamical system is “equivalent” to another becomes, via symbolic dynamics, a coding problem. Roughly speaking, two dynamical systems, (X_1, T_1) and (X_2, T_2) , are *equivalent* if there is an invertible mapping from X_1 to X_2 which makes T_1 “look like” T_2 . If the corresponding symbolic dynamical systems are denoted (Y_1, σ) and (Y_2, σ) , then an equivalence between (X_1, T_1) and (X_2, T_2) becomes a time-invariant, invertible encoding from Y_1 to Y_2 (time-invariant because the shift transformation represents the dynamics). Thus, the classification

problem in dynamical systems leads to a coding problem between constrained sets of sequences.

We have described dynamical systems as the discrete-time iteration of a single mapping. However, continuous-time iterations have been studied since the inception of the subject. These are known as continuous-time flows, with the main example being the set of solutions to a system of ordinary differential equations. Indeed, the work of Hedlund and Morse mentioned above was done in this context.

Shift Spaces and Sliding Block Codes

Let $\mathcal{A}$ be an alphabet of symbols, which we assume to be finite. The principal objects of study in symbolic dynamics are certain kinds of collections of sequences of symbols from $\mathcal{A}$. Typically, these sequences are infinite $x = x_0x_1x_2\ldots$, but it is often more convenient to deal with *bi-infinite sequences* $x = \ldots x_{-2}x_{-1}x_0x_1x_2\ldots$. For some problems, the results are similar in the infinite and bi-infinite categories, while for other problems, they are quite different. In this article, we focus on the bi-infinite setting.

The symbol x_i is the i th *coordinate* of x . When writing a specific sequence, we need to specify which is the 0th coordinate. As suggested in section “Origins of Symbolic Dynamics: Modeling of Dynamical Systems,” this is done with a decimal point to separate the x_i with $i \geq 0$ from those with $i < 0$: $x = \ldots x_{-2}x_{-1}.x_0x_1x_2\ldots$. A *block* or *word* over $\mathcal{A}$ is a finite sequence of symbols from $\mathcal{A}$. A block of length N is called an N -*block*. For blocks u, v , the block uv is the concatenation of u and v , and for a block w , the concatenation of N copies of w is denoted w^N .

The *full $\mathcal{A}$ -shift* $\mathcal{A}^{\mathbb{Z}}$ is the set of all bi-infinite sequences of symbols from $\mathcal{A}$. The *full r -shift* is the full shift over the alphabet $\{0, 1, \ldots, r-1\}$. The *shift map* σ on a full shift maps a point x to the point $y = \sigma(x)$ whose i th coordinate is $y_i = x_{i+1}$.

The *orbit* of a point in a full shift is its orbit under the shift map. The full shift contains many different types of orbits. For instance, it contains a dense orbit (namely, any sequence which contains every block in the alphabet) and periodic orbits (namely, any sequence which is periodic).

We are interested in sets that can be specified by a list (finite or infinite) of forbidden blocks. Namely, given a collection $\mathcal{F}$ of “forbidden blocks” over $\mathcal{A}$, the subset X consisting of all sequences in $\mathcal{A}^{\mathbb{Z}}$, none of whose subwords belong to $\mathcal{F}$, is called a *shift space* (or simply *shift*), and we write $X = X_{\mathcal{F}}$. When a shift space X is contained in a shift space Y , we say that X is a *subshift* of Y .

Example 1 X is the set of all binary sequences with no two 1’s next to each other. Here $X = X_{\mathcal{F}}$, where $\mathcal{F} = \{11\}$. This shift is called the *golden mean shift*, for reasons that will become apparent later.

Example 2 X is the set of all binary sequences so that between any two 1’s there are an even number of 0’s. We can take for $\mathcal{F}$ the collection

$$\{10^{2n+1}1 : n \geq 0\}.$$

This example is naturally called the *even shift*.

Example 3 X is the set of all binary sequences such that between any two successive 1’s, number of 0’s is prime. We can take for $\mathcal{F}$ the collection

$$\{10^n1 : n \text{ is composite}\}.$$

This example is naturally called the *prime shift*.

Alternatively (and equivalently), shift spaces can be defined as closed, shift-invariant subsets of full shifts. Here, “closed” means with respect to a metric, for which two points are close if they agree in a large “central block”; one such metric is $\rho(x, y) = 2^{-k}$ if $x \neq y$, with k maximal such that $x_{[-k, k]} = y_{[-k, k]}$ (with the conventions that $\rho(x, y) = 0$ if $x = y$ and $\rho(x, y) = 2$ if $x_0 \neq y_0$).

Let X be a subset of a full shift, and let $\mathcal{B}_N(X)$ denote the set of all N -blocks that occur in elements of X . The *language of X* is $\mathcal{B}(X) = \cup_N \mathcal{B}_N(X)$. It can be shown that the language of a shift space determines the shift space uniquely, and so we can equally well describe a shift space by specifying the “occurring” or “allowed” blocks, rather than the forbidden blocks. For

example, the golden mean shift is specified by the language of blocks in which 1's are isolated.

This establishes a connection with automata theory (Aho et al. 1974; Béal 1993), which studies collections of blocks, rather than infinite or bi-infinite sequences. The languages that occur in symbolic dynamics, i.e., as $\mathcal{B}(X)$ for some shift space X , are simply those sets $\mathcal{L}$ of blocks that satisfy two simple properties: every sub-block of an element of $\mathcal{L}$ belongs to $\mathcal{L}$, and every element $w \in \mathcal{L}$ is extendable to a larger block $awb \in \mathcal{L}$, with $a, b \in \mathcal{A}$.

Some examples are best described by allowed blocks. One example is the famous *Morse shift*.

Example 4 Let $A_0 = 0$ and inductively define blocks $A_{n+1} = A_n \overline{A_n}$, where $\overline{A_n}$ denotes bitwise complement. The shift space whose allowed blocks are the sub-blocks of the A_n is called the *Morse shift*.

This shift space has an alternative description as follows. Since each A_n is a prefix of A_{n+1} , the sequence of blocks A_n determines a unique right-infinite sequence x^+ (with each A_n as a prefix). If we denote x^- as the left-infinite sequence obtained by writing x^+ backwards, then the Morse shift is the closure of the orbit of the point $x^- \cdot x^+$. The sequence x^+ is known as the Prouhet-Thue-Morse (PTM) sequence, since Prouhet and Thue introduced it earlier, but for different purposes.

While it is not immediately obvious, it can be shown that the PTM sequence is not periodic; moreover, the Morse shift is a *minimal shift*, which means that all its orbits are dense (Morse 1921). In contrast, while the full, golden mean, even and prime shifts all have dense orbits, each also has a dense set of periodic orbits.

There are two simple, but very important, constructions in symbolic dynamics that construct from a given shift space a new version which in some sense looks deeper into the space at the cost of a larger alphabet and more complex description.

Let X be a shift space over the alphabet $\mathcal{A}$, and $\mathcal{A}^{(N)} = \mathcal{B}_N(X)$. We can consider $\mathcal{A}^{(N)}$ as an alphabet in its own right, and form the full shift $(\mathcal{A}^{(N)})^{\mathbb{Z}}$. Define the *Nth higher block code* by

$$(\beta_N(x))_i = x_{[i, i+N-1]} .$$

Then the *Nth higher block shift* or *higher block presentation* of a shift space X is the image $X^{[N]} = \beta_N(X)$ in the full shift over $\mathcal{A}^{(N)}$.

Similarly, define the *Nth higher power code* $\gamma_N : X \rightarrow (\mathcal{A}^{(N)})^{\mathbb{Z}}$ by

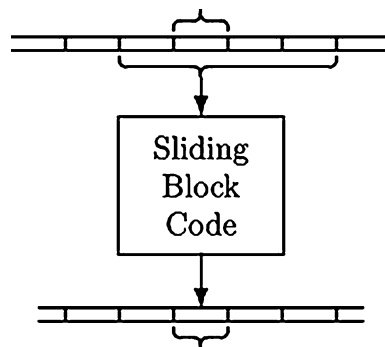
$$(\gamma_N(x))_i = x_{[iN, iN+N-1]} .$$

The *Nth higher power shift* X^N of X is the image $X^N = \gamma_N(X)$ of X . The difference between $X^{[N]}$ and X^N is that the former is constructed by considering overlapping blocks and the latter by non-overlapping blocks.

Next, we turn to mappings between shift spaces. Suppose that $x = \dots x_{-1} \cdot x_0 x_1 \dots$ is a sequence in a shift space X over $\mathcal{A}$. We can transform x into a new sequence $y = \dots y_{-1} \cdot y_0 y_1 \dots$ over another alphabet $\mathcal{C}$ as follows. Fix integers m and n with $-m \leq n$. To compute the i th coordinate y_i of the transformed sequence, we use a function Φ that depends on the “window” of coordinates of x from position $i - m$ to position $i + n$. Here $\Phi : \mathcal{B}_{m+n+1}(X) \rightarrow \mathcal{C}$ is a fixed *block map*, called an $(m + n + 1)$ -*block map* from allowed $(m + n + 1)$ -blocks in X to symbols in $\mathcal{C}$, and so

$$y_i = \Phi(x_{i-m} x_{i-m+1} \dots x_{i+n}) = \Phi(x_{[i-m, i+n]}). .$$

This is illustrated in Fig. 2.



Symbolic Dynamics, Fig. 2 Sliding block code

Let X be a shift space over $\mathcal{A}$, and $\Phi : \mathcal{B}_{m+n+1}(X) \rightarrow \mathcal{C}$ be a block map. Then the map $\phi : X \rightarrow \mathcal{C}^{\mathbb{Z}}$ defined by $y = \phi(x)$, with y_i given by Φ above, is called the *sliding block code* with *memory* m and *anticipation* n induced by Φ . We will denote the formation of ϕ from Φ by $\phi = \Phi_{\infty}^{[-m,n]}$, or more simply by $\phi = \Phi_{\infty}$ if the memory and anticipation of ϕ are understood. If not specified, the memory is taken to be 0. If Y is a shift space contained in $\mathcal{C}^{\mathbb{Z}}$ and $\phi(X) \subseteq Y$, we write $\phi : X \rightarrow Y$.

In analogy with the characterization of shift spaces as closed shift-invariant sets, sliding block codes can be characterized in a topological manner: namely, as the maps between shift spaces that are continuous and commute with the shift. This result is known as the Curtis-Hedlund-Lyndon theorem (Hedlund 1969).

Example 5 Let $\mathcal{A} = \{0, 1\} = \mathcal{C}$, $X = \mathcal{A}^{\mathbb{Z}}$, $m = 0$, $n = 1$, and $\Phi(a_0 a_1) = a_0 + a_1 \pmod{2}$. Let $\phi = \Phi_{\infty} : X \rightarrow X$.

Example 6 The sliding block code, generated by $\Phi(00) = 1$, $\Phi(01) = 0 = \Phi(10)$, maps the golden mean shift onto the even shift.

Example 7 There is a trivial sliding block code from the full 2-shift into the full 3-shift, generated by $\Phi(0) = 0$, $\Phi(1) = 1$.

If a sliding block code $\phi : X \rightarrow Y$ is onto, then ϕ is called a *factor code* or *factor map*, and Y is a *factor* of X . If $\phi : X \rightarrow Y$ is one-to-one, then ϕ is called an *embedding* of X into Y . The sliding block code in Example 7 is an embedding but not a factor code, while the codes in Examples 5 and 6 are factor maps, but not embeddings.

A major (and unrealistic) goal of symbolic dynamics is to classify in an explicit way shift spaces up to the following natural notion of equivalence. A sliding block code $\phi : X \rightarrow Y$ is a *conjugacy* (or *topological conjugacy*) if it is invertible with sliding block inverse. Equivalently, a conjugacy is a bijective sliding block code and therefore simultaneously a factor code and an embedding. If there is a conjugacy from

one shift space X to another Y , we say that X and Y are *conjugate*, denoted $X \cong Y$.

As an example, the higher block map β_N is a conjugacy between a shift space X and its higher block shift $X^{[N]}$. Via this code, we can “re-code” any sliding block code as a 1-block code (though typically a conjugacy and its inverse cannot, by this artifice, be simultaneously re-coded to 1-block codes).

In this section, we have given examples of relatively simple sliding block codes. But the typical conjugacy, as well as factor code and embedding, can be much more complicated.

Shifts of Finite Type and Sofic Shifts

A *shift of finite type* (SFT) is a shift space that can be described by a finite set of forbidden blocks, i.e., a shift space X having the form $X_{\mathcal{F}}$ for some finite set $\mathcal{F}$ of blocks. The terminology shift of finite type (or *subshift of finite type*) comes from dynamical systems (Smale 1967).

An SFT is *M-step* (or has *memory M*) if it can be described by a collection of forbidden blocks all of which have length $M + 1$. It is easy to see that any SFT is *M-step* for some M .

Since any shift space can be defined by many different collections of forbidden blocks, it is useful to have the following equivalent condition expressed in terms of allowed blocks: an SFT is *M-step* if and only if whenever u is an allowed block of length at least M , u' is the suffix of u with length M and a is a symbol, then ua is allowed if and only if $u'a$ is allowed. In other words, in order to tell whether a symbol can be allowably concatenated to the end of an allowed word u , one need only look at the last M symbols of u . This is analogous to the “finite memory” property of *M-step* Markov chains.

The golden mean shift X is a 1-step SFT, since it was defined by a forbidden list consisting of exactly one block: $\mathcal{F} = \{11\}$. Equivalently, it is only the last symbol of an allowed block that determines whether a given symbol can be concatenated at the end. In contrast, the even shift is not an SFT: for any M , the symbol 1 can

be concatenated to the end of exactly one of the (allowed) words 10^M and 10^{M+1} .

Recall that the higher block code β_M is a conjugacy from X to $X^{[M]}$. Via this code, any SFT can be recoded to a 1-step SFT. And so, any sliding block code on a shift space can be recoded to a 1-block code on a 1-step SFT. It is useful to have a concrete description of 1-step SFT's. In fact, these are precisely the shift spaces consisting of all bi-infinite sequences of vertices along paths on a finite directed graph. These are called *vertex shifts*. We find it more convenient to work with sequences of edges instead. To be precise:

Let G be a finite directed graph (or simply graph) with vertices (or *states*) $\mathcal{V} = \mathcal{V}(G)$ and edges $\mathcal{E} = \mathcal{E}(G)$. For an edge e , $i(e)$ denotes the initial state and $t(e)$ the terminal state. A *path* in G is a finite sequence of edges in G such that the terminal state of an edge coincides with the initial state of the following edge; a *cycle* in G is path that begins and ends at the same state. We will assume that G is *essential*, i.e., that every state has at least one outgoing edge and one incoming edge.

The *adjacency matrix* $A = A(G)$ is the matrix indexed by $\mathcal{V}$ with A_{IJ} equal to the number of edges in G with initial state I and terminal state J . Since a graph and its adjacency matrix essentially determine the same information, we will frequently associate a graph G with its adjacency matrix A and a nonnegative integer matrix A with a graph G .

The *edge shift* X_G or X_A is the shift space over the alphabet $\mathcal{A} = \mathcal{E}$ defined by

$$\begin{aligned} X_G &= X_A \\ &= \{ \xi = (\xi_i)_{i \in \mathbb{Z}} \in \mathcal{E}^{\mathbb{Z}} : \text{each } \xi_{i+1} \text{ follows } \xi_i \} . \end{aligned}$$

It can be readily verified that edge shifts are 1-step SFT's. While edge shifts do not include all 1-step SFT's, any 1-step SFT can be recoded to an edge shift, and, compared with vertex shifts, edge shifts offer the advantage of a more compact description.

For many purposes, one can study a general shift space X by breaking it into smaller, more well-behaved pieces. A shift space is *irreducible* if whenever u and w are allowed blocks, there is a

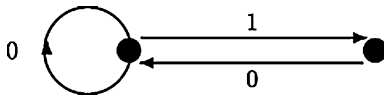
“connecting” block v such that uvw is allowed. While shift spaces do not always decompose into disjoint unions of irreducible shifts, every SFT can be written as a finite disjoint union of irreducible SFT's X_i together with “transient” one-way connections from one X_i to another. And irreducible edge shifts can be characterized in a particularly concrete form: namely, X_G is irreducible if and only if G is *irreducible*, i.e., for every ordered pair of vertices I and J there is a path in G starting at I and ending at J .

There is a stronger notion which is defined by a uniformity condition on the length of the connecting block. A shift space is *mixing* if whenever u and w are allowed blocks, there is an N , possibly depending on u and w , such that for all $n \geq N$, there is block v of length n such that uvw is allowed. And an edge shift X_G is mixing if and only if G is *primitive*, i.e., there is an integer N such that for any $n \geq N$ and any ordered pair of vertices I and J , there is a path in G of length n starting at I and terminating at J . It follows that for SFT's in the definition of mixing, the uniform connecting length N can be chosen independent of the allowed blocks u and w .

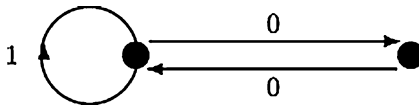
It can be shown that, in some sense, any irreducible SFT X can be broken down into a union of disjoint maximal mixing shifts; namely, X can be written as the disjoint union of finitely many sets X_i , $i = 0, \dots, p-1$ such that $\sigma(X_i) = X_{i+1} \bmod p$ and for each i , σ^p restricted to X_i can be regarded as a mixing SFT. This is a consequence of Perron-Frobenius theory, upon which symbolic dynamics relies heavily; see Seneta (1980) for an introduction to this theory.

A *sofic shift* is the set of bi-infinite sequences obtained from a finite labeled directed graph $\mathcal{G} = (G, \mathcal{L})$; here, G is a finite directed graph and $\mathcal{L}$ is a labeling of the edges of G . The labeled graph is often called a *presentation* of the sofic shift. The golden mean shift and even shift are sofic, with presentations given in Figs. 3 and 4.

SFT's are sofic because any M -step SFT can be presented by a graph whose states are allowed M -blocks. Note also that any sofic shift is a factor of an SFT, namely, via a (1-block) factor code $\mathcal{L}_\infty$ on the edge shift X_G based on a presentation $(G, \mathcal{L})$. In fact, the converse is true, and so, the



Symbolic Dynamics, Fig. 3 Presentation of golden mean shift



Symbolic Dynamics, Fig. 4 Presentation of even shift

sofic shifts are precisely the shift spaces that are factors of SFT's. This was the original definition of sofic shifts given by Weiss (1973).

Typically, the labeling is *right resolving*, which means that at any given state, all outgoing edges have distinct labels (as in Figs. 3 and 4).

Theorem 1.

1. Any sofic shift has a right resolving presentation.
2. Any irreducible sofic shift has a unique minimal right resolving presentation.

Part (a) is a direct consequence of the subset construction in automata theory (Aho et al. 1974; Béal 1993) which constructs a right resolving presentation from an arbitrary presentation; see also Coven and Paul (1975, 1977). Part (b) makes use of the state-minimization algorithm from automata theory (Aho et al. 1974; Béal 1993), but requires an idea beyond that found in automata theory (Fischer 1975a, b). The unique presentation in part (b) may be regarded as a canonical presentation. We remark that for an irreducible (resp. mixing) sofic shift, the underlying graph of the unique minimal right resolving presentation is irreducible (resp., primitive).

Sometimes, it is useful to weaken the concept of right resolving to *right closing*, which means “right resolving with delay”; more precisely a labeling is right closing, with delay D if all paths of length $D + 1$ with the same initial state and the same label have the same initial edge. Also, we sometimes consider *left resolving* and *left closing* labelings (replace “outgoing” with “incoming” in

the definition of right resolving, and replace “initial” with “terminal” in the definition of right closing).

While the class of SFT's is defined by a “finite-memory” property, the more general class of sofic shifts is defined by a “finite-state” property, in that the possible symbols that can occur at time 0 are determined by the past via one of finitely many states. In a presentation, the vertices can be viewed as state information which connects sequences in the past with sequences in the future.

It is not difficult to show that neither the prime shift (Example 3) nor the Morse shift (Example 4) are sofic and therefore also not SFT. For the prime shift, this can be done by an application of the pumping lemma from automata theory (Aho et al. 1974) (or p. 68 of (Lind and Marcus 1995)).

There are uncountably many shift spaces, but only countably many sofic shifts. So, it is not surprising that the behavior of sofic shifts is very special. However, they are very useful in modeling smooth dynamical systems (Section “Origins of Symbolic Dynamics: Modeling of Dynamical Systems”) and in information theory and applications to data recording (Section “Coding for Data Recording Channels”), where they arise as *constrained systems*, although this term is usually reserved for the set of (finite) blocks obtained from a finite directed labeled graph (Marcus et al. 1998).

Entropy and Periodic Points

An invariant of conjugacy is an object associated to a shift space that is preserved under conjugacy. It can be shown that many of the concepts that we have already introduced are invariants: irreducibility, mixing, as well as the properties of being a shift of finite type or sofic shift. Beyond these qualitative invariants, there are many quantitative invariants that can, in many cases, be computed explicitly. Foremost among these is topological entropy.

The (*topological*) *entropy* (or simply *entropy*) of X is:

$$h(X) = \lim_{N \rightarrow \infty} \frac{\log |\mathcal{B}_N(X)|}{N};$$

here, $|\cdot|$ denotes cardinality, and for definiteness the log is taken to mean $\log_2$ but any base will do. A subadditivity argument shows that the limit does indeed exist (Walters 1982). It should be evident that $h(X)$ is a measure of the “size” or “complexity” of X , as it is simply the asymptotic growth rate of the number of blocks that occur in X . Topological entropy for continuous dynamical systems was defined by Adler, Konheim, and McAndrew (1965), in analogy with measure-theoretic entropy.

Since a k -block sliding block code from a shift space X to a shift space Y maps $\mathcal{B}_{N+k-1}(X)$ into $\mathcal{B}_N(Y)$, it is not hard to see that entropy is an invariant of conjugacy and that it cannot increase under factors and cannot decrease under embeddings.

In many cases, one can explicitly compute entropy. For example, for the full r -shift X , $|\mathcal{B}_N(X)| = r^N$, and so $h(X) = \log r$. And from the defining sequence of the Morse shift, we see that the number of distinct 2^N -blocks is at most $4(2^N)$, it follows that the growth rate cannot be exponential and so the entropy of the Morse shift is zero.

For SFT's and more generally for sofic shifts, entropy can be computed explicitly. The key to this computation is the following result, which is based on the Perron-Frobenius theorem.

Theorem 2 Parry (1964), Shannon (1948) For any graph G , $h(X_G) = \log \lambda_{A(G)}$, where $\lambda_{A(G)}$ is the largest eigenvalue of $A(G)$.

The rough idea is that the number of N -blocks in X_G is the number of paths of length N and thus also the sum of the entries of $A(G)^N$, which is controlled by the largest eigenvalue. This is proven first for primitive graphs, whose adjacency matrices have a unique eigenvalue of maximum modulus and this eigenvalue is positive; in fact, for a primitive graph and a pair of states I, J , the number of paths of length N from I to J grows like $\lambda_{A(G)}^N = 2^{Nh(X_G)}$. For general graphs, one uses the

decomposition into primitive, and then irreducible, graphs.

To extend this to a sofic shift Y , one uses a right resolving presentation $(G, \mathcal{L})$ of Y . Since every block of Y is the label of at most $\mathcal{V}(G)$ paths in G , it follows that:

Theorem 3 Let $\mathcal{G} = (G, \mathcal{L})$ be a right-resolving labeled graph presenting a sofic shift Y . Then $h(Y) = h(X_G)$.

From Fig. 3, we see that the golden mean shift is obtained as a right resolving presentation of the graph with adjacency matrix:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

A computation shows that λ_A is the golden mean, and so the entropy of the golden mean shift is the log of the golden mean; this is one explanation of the meaning of the term golden mean shift. From Fig. 4, we see that the even shift has the same entropy.

One cannot overstate the importance of entropy as an invariant. Yet, it is somewhat crude; it is perhaps not surprising that a single numerical invariant would not be sufficient to completely capture the many intricacies of shift spaces, even of sofic shifts or SFT's.

An invariant finer than entropy is the zeta function, which combines information regarding the numbers of periodic sequences of all periods, described as follows.

For a shift space X , let $p_n(X)$ denote the number of points in X of period n (i.e., the number of $x \in X$ such that $\sigma^n(x) = x$). It is straightforward to show that each $p_n(X)$ is an invariant. Since distinct periodic sequences define distinct blocks, it follows that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log p_n(X) \leq h(X).$$

The inequality can be strict; for example, there are shift spaces (such as the direct product of the

Morse shift with the full 2-shift) with positive entropy but no periodic points at all.

However, for irreducible SFT's and sofic shifts, the entropy $h(X)$ can be recovered from the sequence $p_n(X)$. The key to understanding this is the fact that for an edge shift $X = X_A$, $p_n(X) = \text{tr}(A^n)$ and thus is the sum of the n th powers of the (non-zero) eigenvalues of A ; in the case that A is primitive, the largest eigenvalue λ_A strictly dominates the other eigenvalues, and thus for large n ,

$$\log p_n(X) = \log \text{tr}(A^n) \approx n \log \lambda_A \approx nh(X) .$$

This shows that for a mixing SFT, the entropy equals the growth rate of numbers of periodic points. In fact, this result applies to all SFT's and sofic shifts.

Theorem 4 For a sofic shift X ,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log p_n(X) = h(X) .$$

The limsup turns out to be a limit in the case that X is a mixing sofic shift.

Most of what we have stated for $p_n(X)$ here applies equally well to $q_n(X)$, the number of points of *least* period n in X . This follows from the fact that “most” periodic points of period n have least period n .

The periodic point information can be conveniently combined into a single invariant, known as the *zeta function*. For a shift space X ,

$$\zeta_X(t) = \exp \left(\sum_{n=1}^{\infty} \frac{p_n(X)}{n} t^n \right) .$$

For an edge shift X_A , one computes the zeta function to be the reciprocal of a polynomial:

$$\zeta_{X_A}(t) = \frac{1}{t^r \chi_A(t^{-1})} = \frac{1}{\det(I - tA)} ,$$

which is completely determined by the non-zero eigenvalues (with multiplicity) of A (Bowen and

Lanford 1970). As an example, the zeta function of the golden mean shift is $1/(1 - t - t^2)$.

The discussion above applies equally well to SFT's since they are conjugate to edge shifts. For a sofic shift, the zeta function turns out to be a rational function, i.e., quotient of two polynomials. This can be shown by analyzing properties of a right resolving presentation of the sofic shift. From this it turns out that the zeta function of the even shift is $(1 - t)/(1 - t - t^2)$. The technique for computing zeta functions of sofic shifts was developed by Manning (1971) (actually, Manning developed the technique to compute zeta functions of hyperbolic dynamical systems).

So, for sofic shifts, all of the periodic point information is determined by a finite collection of complex numbers, namely, the zeros and poles of the zeta function.

Finally, we mention another simple invariant obtained from the periodic points. The *period*, $\text{per}(X)$, of a shift space X is the gcd of lengths of periodic points in X , i.e., the gcd of the set of n such that $p_n(X) \neq 0$. If $X = X_G$ is an edge shift and G is irreducible, then $\text{per}(X_G) = \text{per}(G)$, which is defined to be the gcd of cycle lengths in G and coincides with the gcd of the lengths of all cycles in G based at any given state. For an irreducible graph with period p and any state I , the number of cycles of length N , a multiple of p , at I grows like $\lambda_{A(G)}^N = 2^{Nh(X_G)}$. Also, an irreducible graph G is primitive if $\text{per}(G) = 1$.

The Conjugacy Problem

The conjugacy problem for SFT's and sofic shifts is a major open problem. After much effort, it remains unsolved today. Much of what we know goes back to R. Williams (1973/74) in the 1970s. One of Williams' main results was that any conjugacy can be decomposed into simple building blocks, as follows.

Let A and B be nonnegative integral matrices, with associated graphs G and H . An *elementary equivalence from A to B* is a pair (R, S) of rectangular nonnegative integral matrices satisfying

$$A = RS, \quad B = SR. \quad (1)$$

In this case, we write $(R, S) : A \equiv B$. A *strong shift equivalence of lag ℓ from A to B* is a sequence of ℓ elementary equivalences

$$\begin{aligned} (R_1, S_1) : A &\equiv A_0 \equiv A_1, \\ (R_2, S_2) : A_1 &\equiv A_2, \\ &\dots, \\ (R_\ell, S_\ell) : A_{\ell-1} &\equiv A_\ell = B. \end{aligned}$$

In this case, we write $A \approx B(\text{lag } \ell)$. We say that A is *strong shift equivalent to B* (and write $A \approx B$) if there is a strong shift equivalence of some lag from A to B .

Via the matrix Eq. (1), one creates a graph K whose state set is the disjoint union of $\mathcal{V}(G)$ and $\mathcal{V}(H)$; for each $I \in \mathcal{V}_G$ and $J \in \mathcal{V}_H$, the graph has R_{IJ} edges from I to J and S_{JI} edges from J to I . The equation $A = RS$ allows one to associate each edge e of G with a unique path $r(e)s(e)$ of length two running from $\mathcal{V}(G)$ to $\mathcal{V}(H)$ to $\mathcal{V}(G)$. Similarly, the equation $B = SR$ allows one to associate each edge e of H with a path $s(e)r(e)$ of length two running from $\mathcal{V}(H)$ to $\mathcal{V}(G)$ to $\mathcal{V}(H)$. One can show that the 2-block sliding block code defined by $\Phi(ef) = s(e)r(f)$ defines a conjugacy from X_A to X_B , and so whenever $A \approx B$, $X_A \cong X_B$. Williams proved this and its converse:

Theorem 5 R. Williams (1973/74) The edge shifts X_A and X_B are conjugate if and only if A and B are strong shift equivalent.

In fact, Williams showed that any conjugacy can be decomposed into a composition of

conjugacies defined by elementary equivalences. This can be interpreted in a way that shows that X_A and X_B are conjugate if and only if we can pass from G to H by a sequence of state splittings and amalgamations, defined as follows.

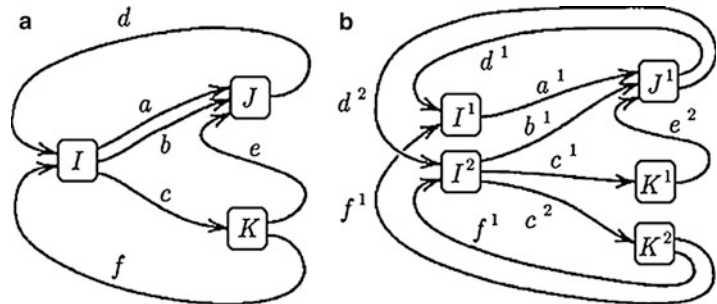
Let G be a graph with states $\mathcal{V}$ and edges $\mathcal{E}$. For each $I \in \mathcal{V}$, partition the outgoing edges from I into disjoint nonempty sets $\mathcal{E}_I^1, \mathcal{E}_I^2, \dots, \mathcal{E}_I^{m(I)}$. Let $\mathcal{P}$ denote the resulting partition of $\mathcal{E}$, and let $\mathcal{P}_I$ denote the partition $\mathcal{P}$ restricted to $\mathcal{E}_I$. The *out-split graph H formed from G using $\mathcal{P}$* has states $I^1, I^2, \dots, I^{m(I)}$, where I ranges over the states in $\mathcal{V}$, and edges e^j , where e is any edge in $\mathcal{E}$ and $1 \leq j \leq m(t(e))$. If $e \in \mathcal{E}$ goes from I to J , then $e \in \mathcal{E}_I^i$ for some i , and we define the e^j to have initial state I^i and terminal state J^j . The 2-block code generated by $\Phi(ef) = e^j$, where $f \in \mathcal{E}_{t(e)}^j$, defines a conjugacy, called an out-splitting code, from X_G to X_H . Similarly, one defines an in-splitting code. Inverses of these conjugacies are called out-amalgamation and in-amalgamation codes.

Figure 5 depicts an out-splitting. The graph (a) on the left has three states I, J, K and the partition elements that define the splitting are $\mathcal{E}_I^1 = \{a\}, \mathcal{E}_I^2 = \{b, c\}, \mathcal{E}_J^1 = \{d\}, \mathcal{E}_J^2 = \{e\}, \mathcal{E}_K^2 = \{f\}$; the graph (b) is the resulting split graph.

By interpreting state splitting and amalgamation in terms of adjacency matrices, one can show that such operations generate elementary equivalences. And one can decompose elementary equivalences into splittings and amalgamations. It follows that:

Theorem 6 R. Williams (1973/74) Every conjugacy from one edge shift to another is the

Symbolic Dynamics,
Fig. 5 A state splitting



composition of splitting codes and amalgamation codes.

This classification for edge shifts naturally extends to SFT's since every SFT is conjugate to an edge shift. It also extends to sofic shifts, and we describe this in the context of irreducible sofic shifts.

Recall that an irreducible sofic shift has a unique minimal right resolving presentation. Any labeled graph can be completely described by a symbolic adjacency matrix, which records the transitions (edges) in the underlying physical graph, as well as the labels of the edges. Namely, the symbolic adjacency matrix is indexed by the states of the underlying graph and the (I, J) -entry is the formal sum of the labels of edges from I to J . It turns out that the notions of elementary equivalence, and hence strong shift equivalence, can be extended to more general categories, in particular to symbolic adjacency matrices.

Theorem 7 Krieger (1984), Nasu (1986) Let X and Y be irreducible sofic shifts. Let A and B be the symbolic adjacency matrices of the minimal right-resolving presentations of X and Y , respectively. Then X and Y are conjugate if and only if A and B are strong shift equivalent.

The classification, provided by these results, would be of limited use if the story ended here. Fortunately, Williams showed that strong shift equivalence yields a strong, delicate, and somewhat computable necessary condition for conjugacy.

Let A and B be nonnegative integral matrices and $\ell \geq 1$. A *shift equivalence* of lag ℓ is a pair (R, S) of rectangular nonnegative integral matrices satisfying the shift equivalence equations

$$AR = RB, \quad SA = BS, \quad A^\ell = RS, \quad B^\ell = SR.$$

We denote this situation by $(R, S) : A \sim B(\text{lag } \ell)$. We say that A is *shift equivalent* to B , written $A \sim B$, if there is a shift equivalence from A to B of some lag. It is not hard to see that an elementary

equivalence is a shift equivalence of lag 1 and that shift equivalence is an equivalence relation. It follows that:

Theorem 8 Williams (1973/74) Strong shift equivalence implies shift equivalence. More precisely, if $A \approx B(\text{lag } \ell)$, then $A \sim B(\text{lag } \ell)$.

Recall from section “Entropy and Periodic Points” that for an edge shift X_A , the set of nonzero eigenvalues, with multiplicity, of A determines the zeta function and hence this set is an invariant of conjugacy. Using the shift equivalence equations, one can show more: the entire Jordan form corresponding to the nonzero eigenvalues is an invariant. This information depends only on properties of the adjacency matrix considered as a linear transformation (over $\mathbf{R}$ or $\mathbf{Q}$). However, A is a nonnegative, integral matrix and both nonnegativity and integrality provide substantially more information. One such invariant that follows from shift equivalence and makes use of integrality is the Bowen-Franks group (Bowen and Franks 1977), $BF(A) = \mathbf{Z}'/\mathbf{Z}'(I - A)$.

Until recently all of the information contained in known conjugacy invariants, such as those above, was subsumed in shift equivalence. And Kim and Roush showed that shift equivalence is decidable (Kim and Roush 1979, 1988), meaning that there is a finite decision procedure via a Turing machine that decides whether two given edge shifts, and therefore two given SFT's, are conjugate (they also showed that a notion of shift equivalence for sofic shifts, formulated by Boyle and Krieger (1986), is decidable (Kim and Roush 1990)). So, a central focus of the subject was the question: is shift equivalence a complete invariant of conjugacy? The answer turns out to be No, as proven by Kim and Roush (1999); see also the survey article Waggoner (1992). However, it is a complete invariant of a weaker form of conjugacy; we say that X_A and X_B are *eventually conjugate* if all sufficiently large powers are conjugate. It is not hard to show that if $A \sim B$, then X_A and X_B are eventually conjugate, and the converse is true as well:

Theorem 9 Kim and Roush (1979), Williams (1973/74) Edge shifts X_A and X_B are eventually conjugate if and only if A and B are shift equivalent.

Also the Kim-Roush counterexamples and subsequent work do not bear on a special case of the conjugacy problem: if A is shift equivalent to the 1×1 matrix $[n]$, is X_A conjugate to the full n -shift? This question, known as the *little shift equivalence problem*, is particularly intriguing because in this case the condition $A \sim [n]$ is simply the statement that A has exactly one non-zero eigenvalue, namely, n .

Shift equivalence can be characterized in another way that has turned out to be very useful. Let A be an $r \times r$ integral matrix. Let R_A denote the real eventual range of A , i.e., $R_A = R^r A^r$. The *dimension group* of A is defined:

$$\Delta_A = \{v \in R_A : vA^k \in \mathbf{Z}^r \text{ for some } k \geq 0\}.$$

The *dimension group automorphism* δ_A of A is the restriction of A to Δ_A , so that $\delta_A(v) = vA$ for $v \in \Delta_A$. The *dimension pair* of A is (Δ_A, δ_A) .

If A is also nonnegative, then we define the *dimension semigroup* of A to be

$$\Delta_A^+ = \{v \in R_A : vA^k \in (\mathbf{Z}^+)^r \text{ for some } k \geq 0\}.$$

The *dimension triple* of A is $(\Delta_A, \Delta_A^+, \delta_A)$.

It can be shown that the dimension triple completely characterizes shift equivalence, i.e., two nonnegative integral matrices are shift equivalent if and only if their dimension groups are isomorphic by an isomorphism that preserves the dimension semigroup and intertwines the dimension group automorphisms. Also, by associating equivalence classes of certain subsets of the shift space X_A to elements of Δ_A^+ , one can interpret the dimension triple in terms of the action of the shift map on X_A (Boyle et al. 1987; Krieger 1980a). And the dimension triple arises prominently in the study of the automorphism group of an SFT, which we now briefly describe. The dimension group for SFT's was developed by Krieger (1980a, b).

In many areas of mathematics, objects are studied by means of their symmetries. This holds true in symbolic dynamics, where symmetries are expressed by automorphisms. An *automorphism* of a shift space X is a conjugacy from X to itself. The set of all automorphisms of a shift space X is a group under composition and is naturally called the *automorphism group*, denoted $\text{aut}(X)$.

The goals are to understand $\text{aut}(X)$ as a group (What kinds of subgroups does it contain? How “big” is it?) and how it acts on X , e.g., given shift-invariant subsets U, V , such as finite sets of periodic points, when is there an automorphism of X that maps U to V ? One might hope that the automorphism group would shed new light on the conjugacy problem for SFT's. Indeed, tools developed to study the automorphism group eventually paved the way for Kim and Roush to find examples of shift equivalent matrices that are not strong shift equivalent. On the other hand, the automorphism group cannot tell the entire story. For instance, $\text{aut}(X_A)$ and $\text{aut}(X_{A^T})$ are isomorphic, since any automorphism read backwards can be viewed as an automorphism of the transposed shift, yet X_A and X_{A^T} may fail to be conjugate (for an example due to Kollmer, see p. 81 of Parry and Tuncel (1982)). It is not even known if the automorphism groups of the full 2-shift and the full 3-shift are isomorphic.

A good deal of our understanding of the action of the automorphism group on an SFT comes from understanding its induced representation as an action of the dimension group; this action is known as the *dimension representation*. For a much more thorough exposition on $\text{aut}(X)$, we refer the reader to Wagoner (1992, 2004).

Other Coding Problems

The difficulties encountered in attempts to solve the conjugacy problem motivated the formulation and study of weaker, but meaningful, notions of equivalence. For instance, we might say that two shift spaces are equivalent if one can be invertibly encoded to the other by some kind of “finite-state machine.” A precise version of this is as follows.

Shift spaces X and Y are *finitely equivalent* if there is an SFT W together with finite-to-one factor codes $\phi_X : W \rightarrow X$ and $\phi_Y : W \rightarrow Y$. We call W a *common extension*, and ϕ_X, ϕ_Y the *legs*.

Here, by “finite-to-one” we mean merely that each point has a finite number of inverse images. It can be shown that any finite-to-one factor code from one shift space to another must preserve entropy, and so entropy is an invariant of finite equivalence. For irreducible sofic shifts, entropy is a complete invariant:

Theorem 10 Parry (1977) Two irreducible sofic shifts are finitely equivalent if and only if they have the same entropy.

Note that from this result and the fact that finite-to-one codes between general shift spaces preserve entropy, for irreducible sofic shifts, we could have just as well-defined finite equivalence with the common extension W merely being a shift space. However, if W is an SFT, we get a more concrete coding interpretation as follows. First, we recode W to an edge shift X_G and recode the legs, $\phi_X = (\Phi_X)_\infty$ and $\phi_Y = (\Phi_Y)_\infty$, to one-block codes, and (with a bit more argument) we can assume that G is irreducible. In this set-up, the finite-to-one condition translates to the so-called “no-diamond” condition, which means that for any given pair of states I, J and finite sequence w , there is at most one path from I to J with label w (Coven and Paul 1975, 1977). Since, for any fixed state I , the number of cycles of length n , a multiple of $p = \text{per}(G)$, at I grows like $2^{nh(W)}$, we have, in this set-up a means to invertibly encode a “large” set of allowed blocks in X to allowed blocks in Y : namely, fix state I , and a large n , which is a multiple of p ; for any cycle γ of length n at state I encode the Φ_X -label of γ to the Φ_Y -label of γ .

For encoding and decoding, one can dispense with state information if the legs are “almost invertible.” A factor code ϕ is *almost invertible* if it is one-to-one on sequences that are typical in the following sense: x is typical if every allowed block appears infinitely often in x both to the left and the right. We then say that shift spaces X and Y are *almost conjugate* if there is an SFT W and

almost invertible factor codes $\phi_X : W \rightarrow X$, $\phi_Y : W \rightarrow Y$. We call (W, ϕ_X, ϕ_Y) an *almost conjugacy* between X and Y .

For irreducible sofic shifts, it can be shown that any almost invertible factor code is finite-to-one, and so almost conjugacy implies finite equivalence. Thus, entropy is again an invariant, and together with a second very mild invariant, it is complete:

Theorem 11 Adler-Marcus (1979) Let X and Y be irreducible sofic shifts with minimal right resolving presentations $(G, \mathcal{L})$ and $(H, \mathcal{M})$. Then X and Y are almost conjugate if and only if $h(X) = h(Y)$ and $\text{per}(G) = \text{per}(H)$.

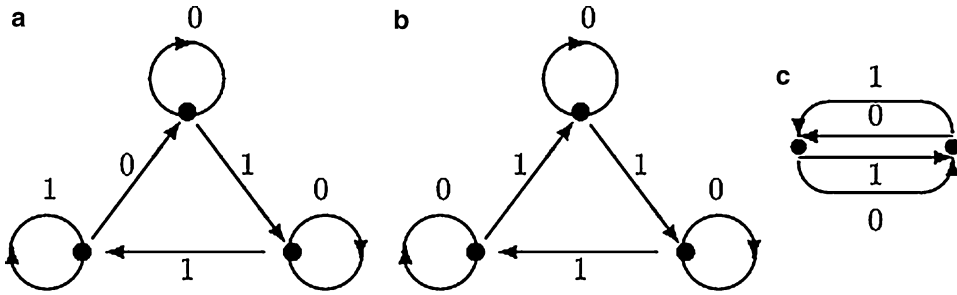
In particular, if X and Y are mixing, then $\text{per}(G) = 1 = \text{per}(H)$ and entropy itself is a complete invariant.

Thus, with an almost conjugacy, one can invertibly encode most sequences in X to those of Y without the need for auxiliary state information. Moreover, if X and Y are almost conjugate, then there is an almost conjugacy of X and Y in which one leg is right-resolving and the other leg is left-resolving (and the common extension is irreducible (Adler and Marcus 1979)). This gives an even more concrete interpretation to the encoding.

The proofs of Theorems 10 and 11 are actually quite constructive. For illustration, we consider a very special, but historically important, case.

Let G be a graph with constant out-degree n . A *road coloring* Φ is a labeling of G such that at each state of G , each symbol $0, \dots, n-1$ appears exactly once as the label of an outgoing edge. An n -ary word w is *synchronizing* if all paths that are labeled w end at the same state. Figure 6 gives examples of road-colorings, with $n = 2$.

For a road-coloring, a binary word may be viewed as a sequence of instructions given to drivers starting at each of the states. A synchronizing word is a word that drives everybody to the same state. For instance, in Fig. 6a, the word 11 drives everybody to the state in the lower-left corner. But Fig. 6b does not have a synchronizing word because whenever a driver takes a “0” road he stays where he is and whenever a driver



Symbolic Dynamics, Fig. 6 Some road-colorings

takes a “1” road he rotates by 120° . The road-coloring in Fig. 6c is essentially the only road-coloring of its underlying graph, and there is no synchronizing word because drivers must always oscillate between the two states.

Now, let X be the full n -shift and Y be an irreducible SFT with entropy $\log n$. Suppose that we could find a presentation $(G, \mathcal{L})$ of Y with G having constant out-degree n and $\mathcal{L}_\infty$ finite-to-one. Then, define Φ to be any road coloring of G . Then Φ_∞ would be a finite-to-one (in fact, right resolving!) factor code from X_G to X . And we would obtain a finite equivalence with $X = X_G$, $\phi_Y = \mathcal{L}_\infty$ and $\phi_X = \Phi_\infty$. If, moreover, we could choose $\mathcal{L}$ and Φ such that ϕ_Y and ϕ_X are almost invertible, then we would have an almost conjugacy.

It turned that this could be arranged for ϕ_Y via a construction related to state splitting (Adler et al. 1977). And if Y were mixing, then G could be chosen to be primitive and ϕ_Y almost invertible. In this setting, $\phi_X = \Phi_\infty$ would be almost invertible iff Φ has a synchronizing word; the sufficiency follows from the fact that every bi-infinite binary sequence which contains w infinitely often to the left would be the label of exactly one bi-infinite sequence of edges (to see this, use the synchronizing and road-coloring properties).

The construction of such a labeling Φ became known as the Road Problem, which remained open for 30 years. In the meantime, a weaker version of the road problem was solved and applied to yield this special case of Theorem 11 (see Adler et al. 1977). Nevertheless, the problem remained an important problem in graph/automata theory and was only recently solved:

Theorem 12 Road Theorem (Trachtman 2007)

If G is a finite directed primitive graph with constant out-degree n , there a road-coloring of G which has a synchronizing word.

Trachtman’s approach relies heavily on earlier work of Friedman (1990) and Kari (2001).

The primitivity assumption above is close to necessary. Clearly some kind of connectivity is required and in the presence of irreducibility, primitivity would be necessary since otherwise there would be a “phase” introduced in the graph that would never allow a word to synchronize, as in Fig. 6c.

So far, we have focused on equivalences between symbolic systems. There has also been considerable attention paid to problems of embedding one system into another and factoring one onto another.

One of the most striking results of this type is the Krieger embedding theorem. It is not hard to show that any proper subshift of an irreducible SFT must have strictly smaller entropy. Thus, a necessary condition for a proper embedding of a shift space into an irreducible SFT is that it have strictly smaller entropy. This condition, together with a trivially necessary condition on periodic points, turns out to be sufficient. Recall that $q_n(X)$ denotes the number of points of least period n in X .

Theorem 13 Embedding Theorem (Krieger 1982)

Let X and Y be irreducible shifts of finite type. Then there is a proper embedding of X into Y if and only if $h(X) < h(Y)$ and for each $n \geq 1$, $q_n(X) \leq q_n(Y)$.

In fact, Krieger's theorem shows that these conditions are necessary and sufficient for a proper embedding of any shift space into a mixing shift space. The analogous problems for embedding into irreducible or mixing sofic shifts are still open, though there are partial results (Boyle 1983).

Using the embedding theorem and other tools from symbolic dynamics, Boyle and Handelman (1991) obtained a stunning application to linear algebra: namely, a complete characterization of the non-zero spectra of primitive matrices over $\mathbf{R}$. In fact, they obtained characterizations of non-zero spectra for primitive matrices over many other subrings of $\mathbf{R}$. While they did not obtain a complete characterization over $\mathbf{Z}$, they formulated a conjecture for $\mathbf{Z}$ and obtained many partial results towards that conjecture, which was later proven using other tools. The result, stated below, shows that three simple necessary conditions on a set of nonzero complex numbers are actually sufficient. In order to state these conditions, we need the following notation:

Let $\Lambda = \{\lambda_1, \dots, \lambda_k\}$ be a list of nonzero complex numbers (with multiplicity). Let $f_\Lambda(t) = \prod_{i=1}^k (t - \lambda_i)$, and $\text{tr}_n(\Lambda) = \sum_{d|n} \mu(n/d) \sum_{i=1}^k \lambda_i^k$, with μ being the Mobius Inversion function.

1. Integrality Condition: $f_\Lambda(t)$ is a monic polynomial (with integer coefficients).
2. Perron Condition: There is a positive entry in Λ , occurring just once, that strictly dominates in absolute value all other entries. We denote this entry by λ_Λ .
3. Net Trace Condition: $\text{tr}_n(\Lambda) \geq 0$ for all $n \geq 1$.

Theorem 14 Kim-Ormes-Roush (2000) Let Λ be a list of nonzero complex numbers satisfying the Integrality, Perron, and Net Trace Conditions. Then there is a primitive integral matrix A for which Λ is the non-zero spectrum of A .

These conditions are all indeed necessary for Λ to be the non-zero spectrum of a primitive integral matrix. The integrality condition is that Λ forms a complete set of algebraic conjugates; the Perron condition states that Λ must satisfy the conditions

of the Perron-Frobenius theorem for primitive matrices; and the Net Trace condition assures that the number of periodic points of least period n would be nonnegative.

We now turn from embeddings to factors. One special case, which is somewhat related to the Road Problem above and also important for data recording applications (Section “Coding for Data Recording Channels”) is:

Theorem 15 Adler et al. (1983a), Marcus (1979) An SFT X factors onto the full n -shift iff $h(X) \geq \log(n)$.

While this special case treats both the equal entropy case ($h(X) = \log(n)$) and unequal entropy case ($h(X) > \log(n)$), in general, the factor problem naturally divides into two cases: lower entropy factors and equal entropy factors. In either case, a trivial necessary condition for Y to be a factor of X is that whenever $q_n(X) \neq 0$, there exists a d/n such that $q_d(Y) \neq 0$. This condition is denoted $P(X) \searrow P(Y)$. Building on ideas from Krieger's embedding theorem, this necessary condition was shown to be sufficient.

Theorem 16 Lower Entropy Factor Theorem (Boyle 1983) Let X and Y be irreducible SFT's with $h(X) > h(Y)$. Then there is a factor code from X to Y if and only if $P(X) \searrow P(Y)$.

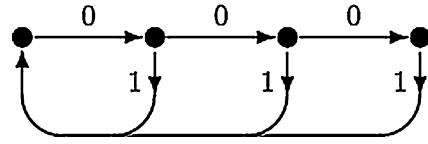
As with the embedding theorem, the lower entropy factors problem for irreducible sofic shifts is still open.

The equal entropy factors problem for SFT's is quite different. Clearly, $P(X) \searrow P(Y)$ is a necessary condition. A second necessary condition involves the dimension group and is simplest to state in the case of mixing edge shifts X_A and X_B .

We say that a subgroup Δ of the dimension group Δ_A is *pure* if whenever an integer multiple of an element $v \in \Delta_A$ is in Δ , then so is v ; intuitively Δ does not have any “rational holes” in Δ_A . The condition is that there is a pure δ_A -invariant subgroup Δ of Δ_A such that (Δ_B, δ_B) is a quotient of $(\Delta, \delta_A|_\Delta)$.

In the equal entropy case, this condition and the trivial periodic point condition, $P(X) \searrow P(Y)$, subsume all known necessary conditions for the

existence of a factor code from one mixing edge shift to another. It is not known if these two conditions are sufficient. For references on this problem, see (Ashley 1991; Boyle et al. 1987; Kitchens et al. 1991).



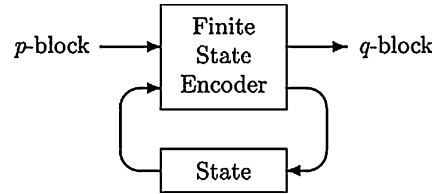
Symbolic Dynamics, Fig. 7 $X(1, 3)$

Coding for Data Recording Channels

In magnetic recording, within any given clock cell, a “1” is represented as a change in magnetic polarity, while a “0” is represented as an absence of such a change. Two successive 1’s (separated by some number, $m \geq 0$, of 0’s) are read as a voltage peak followed by a voltage trough (or vice versa). If the peak and trough occur too close together, *intersymbol interference* can occur: the amplitudes of the peak and trough are degraded, and the positions at which they occur are distorted. In order to control intersymbol interference, it is desirable that 1’s not be too close together, or equivalently that runs of 0’s not be too short. On the other hand, for timing control, it is desirable that runs of 0’s not be too long; this is a consequence of the fact that only 1’s are observed: the length of a run of 0’s is inferred by connection to a clock via a feedback loop, and a long run of 0’s could cause the clock to drift more than one clock cell.

This gives rise to *run-length-limited shift spaces*, $X(d, k)$, where runs of 0’s are constrained to be bounded below by some positive integer d and bounded above by some positive $k \geq d$. More precisely, $X(d, k)$ is defined by the constraints that 1’s occur infinitely often in each direction, and there are at least d 0’s, but no more than k 0’s, between successive 1’s. Note that $X(d, k)$ is an SFT with forbidden list $\mathcal{F} = \{0^{k+1}, 10^i1, 0 \leq i < d\}$. Figure 7 depicts a labeled graph presentation of $X(1, 3)$. The SFT’s $X(1, 3)$, $X(2, 7)$, and $X(2, 10)$ have been used in floppy disks, hard disks, and the compact audio disk, respectively.

Now in order to record completely arbitrary information, we need to build an encoder which encodes arbitrary binary sequences into sequences that satisfy a given constraint (such as $X(d, k)$). The encoder is a finite-state machine, as

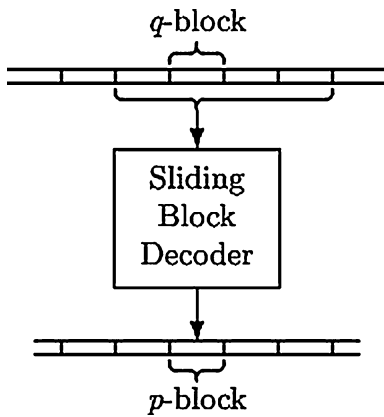


Symbolic Dynamics, Fig. 8 Encoder

depicted in Fig. 8. It maps arbitrary binary data sequences, grouped into blocks of length p (called p -blocks), into constrained sequences, grouped into blocks of length q (called q -blocks). The encoded q -block is a function of the p -block as well as an internal state. When concatenated together, the sequence of encoded q -blocks must satisfy the given constraint. Also, the encoded sequences should be decodable, meaning that given the initial encoder state, a string of p -blocks can be recovered from its encoded string of q -blocks, possibly allowing a fixed delay in time.

If the constrained sequences satisfy the constraints of a sofic shift X , we say that such a code is a *rate $p:q$ finite-state code into X* . In terms of symbolic dynamics, such a code consists of an edge shift X_G , a right resolving factor code ϕ_1 from X_G to the full 2^p -shift and a right closing sliding block code ϕ_2 into X^q (the right closing condition expresses the decodability condition).

In most applications, it is important that a stronger decoding condition be imposed. Namely, a *sliding block decodable rate $p:q$ finite-state code into X* consists of a finite-state code given by (X_G, ϕ_1, ϕ_2) and a sliding block code $\psi : X^q \rightarrow X^p$ such that $\psi \circ \phi_2 = \phi_1$. This means that the decoded p -block depends only upon the local context of the received q -block – that is, decoding is accomplished by applying a time-invariant function to a window consisting of a bounded amount of memory and/or anticipation, but otherwise is state-independent (see Fig. 9, which



Symbolic Dynamics, Fig. 9 Sliding block decoder

depicts the situation where the memory is 1 and the anticipation is 2). The point is that whenever the window of the decoder passes beyond a raw channel error, that error cannot possibly affect future decoding; thus, sliding block decoders control error propagation.

Symbolic dynamics has played an important role in providing a framework for constructing such codes as well as for establishing bounds on various figures of merit for such codes. Specifically, by modifying the constructions used in the proofs of Theorems 10 and 11, in the early 1980s, Adler, Coppersmith, and Hassner (ACH) established the following results:

Theorem 17 Finite-State Coding Theorem (Adler et al. 1983a) Let X be a sofic shift and p, q be positive integers. Then there is a rate $p:q$ finite-state code into X if and only if $p/q \leq h(X)$.

Theorem 18 Sliding Block Decoding Theorem (Adler et al. 1983a) Let X be an SFT and p, q be positive integers. Then there is a rate $p:q$ sliding-block decodable finite-state code into X if and only if $p/q \leq h(X)$.

We have described all of this in the context of the binary alphabet for data sequences. In fact, it works just as well for any finite alphabet, and the method used to prove Theorem 18 solved, at the same time, a special case of the factor problem for SFT's: namely, Theorem 15 above. Sometimes,

Theorems 17 and 18 are stated only for rate 1:1 codes but in the context of arbitrary finite alphabets (e.g., see Chap. 5 in (Lind and Marcus 1995)), this easily extends to the general $p:q$ case by passing to powers.

The ACH paper was the beginning of a rigorous theory of constrained coding. Theorem 18 has been extended to a large class of sofic shifts, and bounds have been established on such figures of merit as number of encoder states as well as the size of the decoding window of the sliding block decoder. While the state-splitting algorithm does construct codes with “relatively small” decoding windows, it is not yet understood how to construct codes with the smallest such windows. This is of substantial engineering interest since the smaller the decoding window, the smaller the error propagation. There is now a substantial literature on the construction of these types of codes, including the state-splitting algorithm as well as many other methods of encoder/decoder design and a wealth of examples that go well beyond the run length limited constraints. See, for example, the expositions (Béal 1993; Immink 2004a; Marcus et al. 1998) as well as the papers (Ashley 1988, 1996; Béal 1990, 2003; Cidecyian et al. 2001; Franaszek 1968, 1982, 1989; Hollmann 1995; Karabed et al. 1999).

Connections with Information Theory and Ergodic Theory

The concept of entropy was developed by Shannon (1948) in information theory in the 1940s and was adapted to ergodic theory in the 1950s and to dynamical systems, and in particular symbolic dynamics, in the 1960s. Shannon focused on entropy for random variables and finite sequences of random variables, but the concept naturally extends to stationary stochastic processes, in particular stationary Markov chains. Roughly speaking, for a stationary process μ , the entropy $h(\mu)$ is the asymptotic growth rate of the number of allowed sequences, weighted by the joint stationary probabilities of μ (for background on entropy and information theory see Cover and Thomas (1991)).

A Markov chain on a graph G is defined by assigning transition probabilities to the edges. Let P denote the stochastic matrix indexed by states of G , with P_{IJ} equal to the sum of the transition probabilities of all edges from I to J . This, together with a stationary vector for P , completely defines the (joint distributions of the) Markov chain.

So, a graph itself can be viewed as specifying only which transitions are possible. For that reason, one could view G as an “intrinsic Markov chain,” and Parry (1964) used this terminology when he introduced SFT’s based on graphs and matrices.

More generally, a stationary process on a shift space X is any stationary process which assigns positive probability only to allowed blocks in X .

For an irreducible graph G with strictly positive transition probabilities on all edges, by Perron-Frobenius theory, P will always have a unique stationary vector. And there is a particular such Markov chain μ_G on G that in some sense distributes probabilities on paths as uniformly as possible. This Markov chain is the most “random” possible stationary process (not just among Markov chains) on X_G in the sense that it has maximal entropy; moreover, its entropy coincides with the topological entropy of X_G .

The Markov chain μ_G is defined as follows: let λ denote the largest eigenvalue of $A(G)$ and w , v denote corresponding left, right eigenvectors, normalized such that $w \cdot v = 1$; for any path γ of length n from state I to state J , we define the stationary probability of γ :

$$\mu_G(\gamma) = \frac{w_I v_J}{\lambda^n}.$$

It is clear from the formula that this distribution is fairly uniform, since all paths with the same initial state, terminal state, and length have the same stationary probability.

This defines a joint stationary distribution on allowed blocks of X_G , and it is not hard to show that it is consistent and Markov, given by assigning transition probability $v_J/(v_I \lambda)$ to each edge from I to J . To summarize:

Theorem 19 Let G be an irreducible graph.

1. Any stationary process μ on G satisfies $h(\mu) \leq \log \lambda$.
2. The Markov chain μ_G is the unique stationary process on G such that $h(\mu_G) = \log \lambda$.

The construction of μ_G is effectively due to Shannon (1948) and uniqueness is due to Parry (1964). The unique entropy-maximizing stationary process on irreducible edge shifts naturally extends to irreducible SFT’s and irreducible sofic shifts (this process will be M -step Markov for an M -step SFT).

Many results in symbolic dynamics were originally proved using μ_G . One example is the fact that any factor code from one irreducible SFT to another of the same entropy must be finite-to-one (Coven and Paul 1974). This result, and many others, were later proven using methods that rely only on the basic combinatorial structure of G and X_G . Nevertheless, μ_G provides much motivation and insight into symbolic dynamics problems, and it illustrates a connection with ergodic theory, which we now discuss.

For background on ergodic theory (such as the concepts of measure-preserving transformations, homomorphisms, isomorphisms, ergodicity, mixing, and measure-theoretic entropy), we refer the reader to Petersen (1989), Rudolph (1990), and Walters (1982). Observe that a stationary process on a shift space X can be viewed as a measure-preserving transformation: the transformation is the shift mapping and the measure on “cylinder sets,” consisting of $x \in X$ with prescribed coordinate values $x_i = a_i, \dots, x_j = a_j$, is defined as the probability of the word $a_i \dots a_j$; the stationarity of the process translates directly into preservation of the measure. The symbolic dynamical concepts of irreducibility and mixing correspond naturally to the concepts of ergodicity and (measure-theoretic) mixing in ergodic theory.

It is well-known (Petersen 1989; Walters 1982) that the measure-preserving transformation (MPT) defined by a stationary Markov chain μ on an irreducible (resp., primitive) graph G is ergodic (resp., mixing) if μ assigns strictly positive conditional probabilities to all edges of G .

Now, suppose that G and H are irreducible graphs and $\phi : X_G \rightarrow X_H$ is a factor code. For a

stationary measure μ on G , we define a stationary measure $\nu = \phi(\mu)$ on X_H by transporting μ to X_H : for a measurable set A in X_H , define

$$\nu(A) = \mu(\phi^{-1}(A)) .$$

Then ϕ defines a measure-preserving homomorphism from the MPT defined by μ to the MPT defined by ν . Since measure-preserving homomorphisms between MPT's cannot reduce measure-theoretic entropy, we have $h(\nu) \leq h(\mu)$.

Suppose that now ϕ is actually a conjugacy. Then it defines a measure-preserving isomorphism, and so $h(\mu) = h(\nu)$. If $\mu = \mu_G$, then by uniqueness, we have $\nu = \mu_H$. Thus, ϕ defines a measure-theoretic isomorphism between the MPT defined by μ_G and the MPT defined by μ_H . In fact, this holds whenever ϕ is merely an almost invertible factor code. This establishes the following result:

Theorem 20 Let G, H be irreducible graphs, and let μ_G, μ_H be the stationary Markov chains of maximal entropy on G, H . If X_G, X_H are almost conjugate (in particular, if they are conjugate), then the measure-preserving transformations defined by μ_G and μ_H are isomorphic.

Hence, conjugacies and almost conjugacies yield isomorphisms between measure-preserving transformations defined by stationary Markov chains of maximal entropy. In fact, the isomorphisms obtained in this way have some very desirable properties compared to the run-of-the-mill isomorphism. For instance, an isomorphism ϕ between stationary processes typically has an infinite window; i.e., to know $\phi(x)_0$, you typically need to know all of x , not just a central block $x_{[-n, n]}$ (these are the kinds of isomorphisms that appear in general ergodic theory and in particular in Ornstein's celebrated isomorphism theory (Ornstein 1970)). In contrast, by definition, a conjugacy always has a finite window of uniform size. It turns out that an isomorphism obtained from an almost conjugacy, as well as its inverse, has finite expected coding length in the sense that to know $\phi(x)_0$, you need to know only a central

block $x_{[-n(x), n(x)]}$, where the function $n(x)$ has finite expectation (Adler and Marcus 1979). In particular, by Theorem 11, whenever X_G and X_H are mixing edge shifts with the same entropy, the measure-preserving transformations defined by μ_G and μ_H are isomorphic via an isomorphism with finite expected coding length.

The notions of conjugacy, finite equivalence, almost conjugacy, embedding, factor code, and so on can all be generalized to the context of stationary measures, in particular to stationary Markov chains. For instance, a conjugacy between two stationary measures is a map that is simultaneously a conjugacy of the underlying shift spaces and an isomorphism of the associated measure-preserving transformations. Many results in symbolic dynamics have been generalized to the context of stationary Markov chains. There is a substantial literature on this, in particular on finitary isomorphisms with finite expected coding time, e.g., (Krieger 1983; Marcus and Tuncel 1990; Mouat and Tuncel 2002; Parry 1979; Schmidt 1984). The expositions (Parry 1991; Parry and Tuncel 1982) give a nice introduction to the subject of strong finitary codings between stationary Markov chains. See also the research papers (Gomez 2003; Marcus and Tuncel 1991, 1993; Parry and Schmidt 1984; Parry and Tuncel 1981; Tuncel 1981, 1983).

Higher Dimensional Shift Spaces

In this section, we introduce higher dimensional shift spaces. For a more thorough introduction, we refer the reader to Lind (2004). For the related subject of tiling systems, see Robinson (2004), Radin (1996), and Mozes (1989, 1992).

The d -dimensional full $\mathcal{A}$ -shift is defined to be $\mathcal{A}^{\mathbb{Z}^d}$. Ordinarily, $\mathcal{A}$ is a finite alphabet, and here we restrict ourselves to this case. An element x of the full shift may be regarded as a function $x : \mathbb{Z}^d \rightarrow \mathcal{A}$, or, more informally, as a “configuration” of alphabet choices at the sites of the integer lattice $\mathbb{Z}^d$.

For $x \in \mathcal{A}^{\mathbb{Z}^d}$ and $F \subseteq \mathbb{Z}^d$, let x_F denote the restriction of x to F . The usual metric on the one-dimensional full shift naturally generalizes to a

metric on $\mathcal{A}^{\mathbf{Z}^d}$ given by $\rho(x, y) = 2^{-k}$, where k is the largest integer such that $x_{[-k, k]^d} = y_{[-k, k]^d}$ (with the usual conventions when $x = y$ and $x_0 \neq y_0$). In analogy with one dimension, according to this definition, two points are “close” if they agree on a large cube $[-k, k]^d$.

We define higher dimensional shift spaces, with the following terminology. A *shape* is a finite subset F of $\mathbf{Z}^d$, and a *pattern* f on a shape F is a function $f : F \rightarrow \mathcal{A}$. We say that X is a *d-dimensional shift space* (or *d-dimensional shift*) if it can be represented by a list $\mathcal{F}$ (finite or infinite) of “forbidden” patterns

$$X = X_{\mathcal{F}} = \{x \in \mathcal{A}^{\mathbf{Z}^d} : \sigma^n(x)_F \notin \mathcal{F} \text{ for all } n \in \mathbf{Z}^d \text{ and all shapes } F\}.$$

Just as in one dimension, we can equivalently define a shift space to be a closed (with respect to the metric ρ) translation-invariant subset of $\mathcal{A}^{\mathbf{Z}^d}$. Here “translation-invariance” means that $\sigma^n(X) = X$ for all $n \in \mathbf{Z}^d$, where σ^n is the translation in direction n defined by $(\sigma^n(x))_m = x_{m+n}$.

We say that a pattern f on a shape F *occurs* in a shift space X if there is an $x \in X$ such that $x_F = f$. Hence the analogue of the language of a shift space is the set of all occurring patterns.

A *d-dimensional shift of finite type* X is a subset of $\mathcal{A}^{\mathbf{Z}^d}$ defined by a finite list $\mathcal{F}$ of forbidden patterns. Just as in one dimension, a *d-dimensional shift of finite type* X can also be defined by specifying allowed patterns instead of forbidden patterns, and there is no loss of generality in requiring the shapes of the patterns to be the same. Thus, we can specify a finite list $\mathcal{L}$ of patterns on a fixed shape F , and set

$$X = X_{\mathcal{L}^c} = \left\{ x \in \mathcal{A}^{\mathbf{Z}^d} : \text{for all } n \in \mathbf{Z}^d, \sigma^n(x)_F \in \mathcal{L} \right\}.$$

In fact, there is no loss in generality in assuming that F is a *d-dimensional cube* $F = [0, k]^d$.

Given a finite list $\mathcal{L}$ of patterns on a shape F , we say that a pattern f on a shape F' is *$\mathcal{L}$ -admissible* (in the shift of finite type $X = X_{\mathcal{L}^c}$) if each of its sub-patterns, whose shape is a

translate of F , belongs to $\mathcal{L}$. Of course, any pattern which occurs in X is $\mathcal{L}$ -admissible. But an $\mathcal{L}$ -admissible pattern need not occur in X .

The analogue of vertex shift (or 1-step shift of finite type) in higher dimensions is defined by a collection of d transition matrices $A_1, \dots, A_d$ all indexed by the same set of symbols $\mathcal{A} = \{1, \dots, m\}$. We set

$$\Omega(A_1, \dots, A_d) = \left\{ x \in \{1, \dots, m\}^{\mathbf{Z}^d} : A_i(x_n, x_{n+e_i}) = 1 \text{ for all } n, i \right\},$$

where e_i is as usual the i th standard basis vector and $A_i(a, b)$ denotes the (a, b) -entry of A_i . Such a shift space is called a *matrix subshift*. When $d = 2$, this amounts to a pair of transition matrices A_1 and A_2 with identical vertex sets. The matrix A_1 controls transitions in the horizontal direction and the matrix A_2 controls transitions in the vertical direction. Note that any matrix subshift is a shift of finite type and, in particular, can be specified by a list $\mathcal{L}$ of patterns on the unit cube $F = \{(a_1, \dots, a_n) : a_i \in \{0, 1\}\}$; specifically, $\Omega(A_1, \dots, A_n) = X_{\mathcal{L}^c}$ where $\mathcal{L}$ is the set of all patterns $f : F \rightarrow \{1, \dots, m\}$ such that if $n, n + e_i \in F$, then $A_i(f(n), f(n + e_i)) = 1$. When we speak of admissible patterns for a matrix subshift, we mean $\mathcal{L}$ -admissible patterns with this particular $\mathcal{L}$. Just as in one dimension, we can recode any shift of finite type to a matrix subshift.

Higher dimensional SFT's can behave very differently from one dimension. For example, there is a simple method to determine if a one-dimensional edge shift, and therefore, a one-dimensional shift of finite type, is nonempty, and there is an algorithm to tell, for a given finite list $\mathcal{L}$, whether a given block occurs in $X = X_{\mathcal{L}^c}$. The corresponding problems in higher dimensions, called the *nonemptiness problem* and the *extension problem*, turn out to be undecidable (Berger 1966; Robinson 1971); see also Kitchens and Schmidt (1988). Even for two-dimensional matrix subshifts X , these decision problems are undecidable. On the other hand, there are some special classes where these problems are decidable. This class includes any two-dimensional matrix subshift such that A_1 commutes with A_2 and $A_2^\top$. For this class, any admissible pattern on a

cube must occur, and so the nonemptiness and extension problems are decidable; see Markley and Paul (1981a, b).

A point x in a d -dimensional shift X is *periodic* if its orbit $\{\sigma^n(x) : n \in \mathbf{Z}^d\}$ is finite. Observe that this reduces to the usual notion of periodic point in one dimension. Now an ordinary (one-dimensional) nonempty shift of finite type is conjugate to an edge shift X_G , where G has at least one cycle. Hence a one-dimensional shift of finite type is nonempty if and only if it has a periodic point. This turns out to be false in higher dimensions (Berger 1966; Robinson 1971), and this fact is intimately related to the undecidability results mentioned above. While one can formulate a notion of zeta function for keeping track of numbers of periodic points, the zeta function is hard to compute, even for very special and explicit matrix subshifts, and, even in this setting, it is not a rational function (Lind 1996).

In higher dimensions, the entropy of a shift is defined as the asymptotic growth rate of the number of occurring patterns in arbitrarily large cubes. In particular, for two-dimensional shifts it is defined by

$$h(X) = \lim_{n \rightarrow \infty} \frac{1}{n^2} \log |X_{[0,n-1] \times [0,n-1]}|,$$

where $X_{[0,n-1] \times [0,n-1]}$ denotes the set of occurring patterns on the square

$$[0, n-1] \times [0, n-1]$$

that occur in X . Recall from section “Entropy and Periodic Points” that it is easy to compute the entropy of a (one-dimensional) shift of finite type using linear algebra. But in higher dimensions, there is no analogous formula and, in fact, other than the group shifts mentioned below, the entropies of only a very few higher dimensional shifts of finite type have been computed explicitly. Even for the two-dimensional “golden mean” matrix subshift defined by the horizontal and vertical transition matrices

$$A_1 = A_2 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

an explicit formula for the entropy is not known. However, there are good numerical approximations to the entropy of some matrix subshifts (e.g., Calkin and Wilf 1998; Nagy and Zeger 2000). And recently, the set of numbers that can occur as entropies of SFT’s in higher dimensions has been characterized (Hochman 2009; Hochman and Meyerovitch 2007); this characterization turns out to be remarkably different from the analogous characterization in one dimension (Lind 1984).

One of the few two-dimensional SFT’s for which entropy has been computed is the domino tiling system (see (Kastelyn 1961), Chap. 5 in (Schmidt 1990)), which consists of all possible tilings of the plane using the 1×2 and 2×1 dominoes. This can be translated into an SFT with four symbols L, R, T, B , subject to the constraints:

$$x_{i,j} = L \Rightarrow x_{i+1,j} = R,$$

$$x_{i,j} = R \Rightarrow x_{i-1,j} = L,$$

$$x_{i,j} = T \Rightarrow x_{i,j-1} = B,$$

$$x_{i,j} = B \Rightarrow x_{i,j+1} = T.$$

The entropy of this SFT is given by a remarkable integral formula:

$$\frac{1}{4} \int_0^1 \int_0^1 (4 - 2 \cos(2\pi s) - 2 \cos(2\pi t)) ds dt$$

Further work along these lines can be found in Kenyon (2008) and Schwartz and Bruck (2008).

One can formulate notions of irreducibility and mixing for higher dimensional shift spaces. It turns out that for SFT’s there are several notions of mixing that all coincide in one dimension, but are vastly different in higher dimensions. For instance, a higher dimensional shift space is *strongly irreducible* if there is an integer $R > 0$ such that for any two shapes F, F' of distance at

least R , any occurring configurations on F and F' can be combined to form an occurring configuration on $F \cup F'$ (Burton and Steif 1994; Ward 1994). For one-dimensional SFT's, this is equivalent to mixing, but much stronger than mixing for two-dimensional SFT's.

Just as in one dimension, we have the notion of sliding block code for higher dimensional shifts. For finite alphabets $\mathcal{A}, \mathcal{B}$, a cube $F \subset \mathbb{Z}^d$ and a function $\Phi : \mathcal{A}^F \rightarrow \mathcal{B}$, the mapping $\phi = \Phi_\infty : \mathcal{A}^{\mathbb{Z}^d} \rightarrow \mathcal{B}^{\mathbb{Z}^d}$ defined by

$$\Phi_\infty(x)_n = \Phi(x_{n+F})$$

is called a sliding block code. By restriction, we have the notion of sliding block code from one d -dimensional shift space to another. As expected, for d -dimensional shifts X and Y , the sliding block codes $\phi : X \rightarrow Y$ coincide exactly with the continuous translation-commuting maps from X to Y , i.e., the maps which are continuous with respect to the metric ρ , defined above, and which satisfy $\phi \circ \sigma^n = \sigma^n \circ \phi$ for all $n \in \mathbb{Z}^d$. Thus, it makes sense to consider the various coding problems, in particular the conjugacy, factor, and embedding problems, in the higher dimensional setting, but these are very difficult.

Even the question of determining when a higher dimensional SFT of entropy at least $\log n$ factors onto the full n -shift seems very difficult (in contrast to Theorem 15). However, there are some positive results for strongly irreducible SFT's. For instance, it is known that any strongly irreducible SFT of entropy strictly larger than $\log n$ factors onto the full n -shift (Desai 2009; Johnson and Madden 2005). In fact, that result requires only a weaker assumption than strong irreducibility, but still much stronger than mixing; recent examples (Boyle et al. 2010) show, among other things, that one cannot weaken that assumption to mere mixing.

There are results on other coding problems as well. For instance, a version of the Embedding theorem in one dimension (Theorem 13) holds for SFT's in two dimensions with a strong mixing property (Lightwood 2003/04); however, it is required that the shift to be embedded contains

no points that are periodic in any single direction. And some other results on entropy of proper subshifts of strongly irreducible SFT's carry over from one dimension to higher dimensions (Pavlov 2011; Quas and Trow 2000). But in many cases where versions of the result carry over, the proofs are much different from those in one dimension.

Measures of maximal entropy for two-dimensional SFT's behave very differently from the one-dimensional case. For instance, even with very strong mixing properties, such as strong irreducibility, there can be more than one measure of maximal entropy (Burton and Steif 1994), and the relationships among entropy-preserving, finite-to-one, and almost invertibility for factor codes discussed in section “Other Coding Problems” can be very different in higher dimensions (Meester and Steif 2001). Other differences with respect to entropy can be found in Quas and Sahin (2003).

There is also the natural notion of higher dimensional sofic shifts, which can be defined as those shift spaces that are factors of SFT's. Recall that every one-dimensional sofic shift is a right-resolving, and hence entropy-preserving, factor of an SFT. It is not known if there is an analogue to this fact in higher dimensions, although recently there has been some progress: every sofic shift Y is a factor of an SFT with entropy arbitrarily close to $h(Y)$ (Desai 2006).

Finally, there is a subclass of d -dimensional SFT's which is somewhat tractable, namely, d -dimensional shifts with group structure in the following sense. Let $\mathcal{A}$ be a (finite) group. Then the full d -dimensional shift over $\mathcal{A}$ is also a group with respect to the coordinate-wise group structure. A (higher-dimensional) *group shift* is a subshift of $\mathcal{A}^{\mathbb{Z}^d}$ which is also a subgroup. For a survey on results for this class, we refer the reader to Lind and Schmidt (2002).

Future Directions

The future directions of the subject will likely be determined by progress on solutions to open problems. In the course of describing topics in this

entry, we have mentioned many open problems along the way. For a much more complete list on a wealth of subareas of symbolic dynamics, we refer the reader to the article (Boyle 2007). While it is difficult to single out the most important challenges, certainly the problem of understanding multidimensional shift spaces, especially of finite type, is one of the most important.

Addendum to the Second Edition

Since the original publication of this article, there has been an enormous amount of activity and progress in symbolic dynamics. While some major problems remain unsolved, with little or no progress, others have been solved, partially or completely, and vast new areas of exploration have been carved out. It would be impossible in this brief update to summarize the developments in symbolic dynamics within the past 10 years. Here, we focus mainly on developments within subareas of symbolic dynamics most closely related to the content of the original article. We have also included some major results known at the time of the original article but not included in the article.

The symbolic modeling approach, introduced in the section on the “[Origins of Symbolic Dynamics: Modelling of Dynamical Systems](#),” continues to be an essential tool in modelling smooth dynamical systems. Markov partitions, such as those for the main examples, hyperbolic toral automorphisms and Smale’s horseshoe, have finitely many “rectangles” corresponding to a finite alphabet, upon which a shift space model is built. One can also consider shift spaces over countably infinite alphabets instead. Recently Markov partitions with countably many rectangles were constructed by Sarig (2013) to model surface diffeomorphisms by shift spaces over countably infinite alphabets. This enabled him to prove a fundamental conjecture of Katok [(2007), problem 7.4] on the growth rate of periodic points for arbitrary surface diffeomorphisms with positive entropy.

The conjugacy problem for shifts of finite type (SFTs), considered in the section “[The Conjugacy](#)

[Problem](#),” is one of the most fundamental problems in the subject: given two SFTs X and Y , how can you tell whether X and Y are conjugate? While many sensitive invariants of conjugacy are known, it is still not known if this problem is even decidable by a finite algorithm.

As described in the “[The Conjugacy Problem](#)” section, conjugacy between a pair of SFTs is equivalent to strong shift equivalence (SSE) of the corresponding adjacency matrices. SSE implies shift equivalence (SE) which is known to be decidable and, for a long time, was conjectured to be equivalent to SSE. But as mentioned in the section, Kim and Roush (1999) showed that this is false even for irreducible matrices. Since then there has been very little progress on the conjugacy problem per se.

SSE and SE are purely algebraic notions for matrices over the semi-ring $\mathbb{Z}^+$. However, they can be defined as meaningful relations over other categories, in particular rings. It is known that over $\mathbb{Z}$, SSE and SE are equivalent. But there are rings over which SSE and SE are not equivalent. Boyle and Schmedling showed that the extent to which they differ is captured by invariants from algebraic K-theory (Boyle and Schmieding 2019).

The section on “[Other Coding Problems](#)” began with notions of equivalence of SFTs weaker than conjugacy, namely, finite equivalence and almost conjugacy, for which one can effectively decide equivalence by simple invariants, namely, entropy and period (Theorems 10 and 11). A version of Theorem 11 has been extended to the setting of a certain class of shift spaces over countably infinite alphabets by Boyle, Buzzi, and Garcia (2006).

Theorems 10 and 11 are intimately related to finite-to-one factor maps (also called finite-to-one factor codes). Such a map has a “degree,” defined as the “typical” number of preimages of any given point. Allahbakhshi and Quas (2013) introduced the analogous notion of class degree for infinite-to-one factor maps. They used this notion to obtain bounds on the number of maximal relative entropy ergodic lifts of a given ergodic measure on the range. The class degree has also turned out to be key to understanding the structure of infinite-to-one factor codes. This was further developed

by Allahbakhshi, Hong, and Jung (2015) who established dynamical properties for infinite-to-one factor maps analogous to those of the fibers of a finite-to-one factor map.

The Embedding Theorem (Theorem 13) and Lower Entropy Factor Theorem (Theorem 16) for irreducible SFTs have been generalized by Thomsen (2004) for special classes of mixing sofic shifts. He also developed structure to explain why his results cannot hold for all mixing sofic shifts. A recent preprint of Krieger (2018) stated a complete necessary, sufficient, and decidable set of conditions for the Lower Entropy Factor Theorem for all mixing sofic shifts.

Lind (1984) characterized the set of numbers that can occur as entropies of SFTs in terms of a set of algebraic integers, known as Perron numbers; specifically, the entropies of SFTs are exactly the logs of roots of Perron numbers. These are positive, in particular real, algebraic integers. Thurston (1990) introduced a generalization known as complex Perron numbers. As shown by Kenyon (1996), these complex numbers are precisely the expansion factors λ for a collection of self-similar tilings of the plane; these are certain tilings by translates of a finite set of basic tiles s.t. for each copy T of a basic tile in the tiling, λT is tiled by translates of the same set of basic tiles. For further work in this area, see Thurston [(2014), especially Theorem 1.3].

The characterization of the collection of zeta functions of SFTs (equivalently, numbers of periodic orbits of all periods) given in Theorem 14 was conjectured by Boyle and Handelman (1991) and proven by Kim, Ormes, and Roush (Kim et al. 2000). In particular, this result characterized the possible multi-sets of non-zero eigenvalues of primitive square matrices over $\mathbb{Z}$. One can ask for similar characterizations if one replaces $\mathbb{Z}$ by other rings. A more general spectral conjecture, which pertains to a wide class of subrings S of $\mathbb{R}$, remains open (Boyle 2007), Conjecture 6.1]. The case $S = \mathbb{R}$ was already obtained in Boyle and Handelman (1991). For more recent work on this subject, see Boyle and Schmieding (2016).

To every SFT, there is an associated continuous suspension flow, and Franks completely classified

irreducible SFTs up to a notion of flow conjugacy by simple invariants (Franks 1984). This classification has been extended to a large class of irreducible sofic shifts by Boyle, Carlsen, and Eilers (2018).

The state-splitting algorithm for constructing sliding block decodable encoders, introduced in the section on “Coding for Data Recording Channels,” was used widely in practice in the 1980s and 1990s. Variants of the algorithm were used to aid in the construction of codes which became standards in the recording industry, including the (1,7) code (Adler et al. 1983b) for hard disk drives, EFMPlus for the DVD (Immink 1994) (see also [(Immink 2004b), section 11.5.2]) and the code used in the first generation of Linear Tape Open (Ashley et al. 1999). Today, the algorithm is used mainly as a proof of concept in new schemes such as codes for radio frequency identification (Barbero et al. 2014), weakly constrained codes (Elishco et al. 2016), and parity-preserving codes (Roth and Siegel 2018).

A major development over the past two decades has been the extension of symbolic dynamics to the multidimensional setting. The set-up was described in the section on “Higher Dimensional Shift Spaces.” Here, bi-infinite sequences are replaced by arrays of symbols on the points of the integer lattice $\mathbb{Z}^d$. As mentioned in that section, Hochman and Meyerovitch (2007) characterized the set of numbers which can occur as the entropy of a $\mathbb{Z}^d$ SFT, $d \geq 2$, as the set of all right recursively enumerable numbers (this is very different from Lind’s characterization in $d = 1$ (Lind 1984)). This result led to a development that has strongly tied symbolic dynamics to computability theory. Many of the central results in this development are summarized in Jeandel (2016).

One central concept in this development is the notion of an effective shift space X , which means a shift space for which there exists a list F of forbidden patterns that defines X and can be produced by a Turing machine. Of course, any $\mathbb{Z}^d$ -SFT or sofic shift is effective. But even for $d = 1$, there are many effective shifts that are not sofic. Nevertheless, any effective $\mathbb{Z}^d$ -shift space X can be “simulated” by a $\mathbb{Z}^{d+1}$ -SFT Y in the sense

that X is a $\mathbb{Z}^d$ -subaction of a factor of Y (Aubrun and Sablik 2013).

Using essentially the same argument to show that the entropy of a $\mathbb{Z}^d$ -SFT is irre, one can show that the entropy of an effective $\mathbb{Z}^d$ -shift space is irre. Since the collection of sofic shift spaces lies in between the SFTs and the effective shift spaces, it follows that the entropies of the sofic shift spaces are exactly the irre numbers. So, for every $\mathbb{Z}^d$ sofic shift there is a $\mathbb{Z}^d$ SFT with the same entropy. As mentioned in the original article, a major open problem is whether for every $\mathbb{Z}^d$ sofic shift X there is a $\mathbb{Z}^d$ SFT which factors onto X and has the same entropy. This problem remains unsolved.

The extension from $\mathbb{Z}$ to $\mathbb{Z}^d$ symbolic dynamics can be carried further. Namely, for an arbitrary countable group G and a finite alphabet $\mathcal{A}$, a G -shift space is a closed subset of the full shift space $\mathcal{A}^G$ that is invariant under the action of G on $\mathcal{A}^G$: for $g \in G$, define $\hat{g}: \mathcal{A}^G \rightarrow \mathcal{A}^G$ by:

$$(\hat{g}(x))_h = x_{hg}$$

Much of the classical theory for Z and Z^d symbolic dynamics has been generalized to large classes of amenable groups G . And more recently a notion of sofic entropy for dynamics of certain non-amenable groups has played a prominent role (Bowen 2012, 2018).

There are many other subareas of symbolic dynamics that are very active and of great importance today. For instance, in algebraic symbolic dynamics one endows the shift space with a group structure such that the shift map is an automorphism. One can characterize many symbolic dynamical properties of such shift spaces in terms of algebraic objects (Schmidt 2012). In cellular automata, one studies the iteration of sliding block codes from one shift space to another as dynamical systems in their own right. These dynamical systems exhibit a rich variety of dynamical behavior (Ceccerini-Silberstein and Coornaert 2010). Thermodynamic formalism studies the notion of topological pressure, a generalization of topological entropy and equilibrium states, and a generalization of measures of maximal entropy (Keller 1998; Sarig 2009). This is

intimately related to the theory of Gibbs states in statistical mechanical systems (Ruelle 2004).

As in the original article, for a list of major open problems in Symbolic Dynamics and progress on their solutions, we refer the reader to Boyle (2007).

Bibliography

- Adler R, Marcus B (1979) Topological entropy and equivalence of dynamical systems. *Memoirs of the American Mathematical Society*, vol 219. AMS, Providence
- Adler R, Weiss B (1970) Similarity of automorphisms of the torus. *Memoirs of the American Mathematical Society*, vol 98. AMS, Providence
- Adler R, Konheim A, McAndrew M (1965) Topological entropy. *Trans Am Math Soc* 114:309–319
- Adler RL, Goodwyn LW, Weiss B (1977) Equivalence of topological Markov shifts. *Isr J Math* 27:48–63
- Adler RL, Coppersmith D, Hassner M (1983a) Algorithms for sliding block codes – an application, of symbolic dynamics to information theory. *Trans IEEE Inf Theory* 29:5–22
- Adler RL, Hassner M, Moussouris JP (1983b, November 1) Method and apparatus for generating a noiseless sliding block code for a (1,7) channel with rate 2/3. United States patent 4,413,251
- Aho AV, Hopcroft JE, Ullman JD (1974) The design and analysis of computer algorithms. Addison-Wesley, Reading
- Allahbakhshi M, Quas A (2013) Class degree and relative maximal entropy. *Trans Am Math Soc* 365:1347–1368
- Allahbakhshi M, Hong S, Jung U (2015) Class closing factor codes and constant-to-one factor codes from shifts of finite type. *Dyn Syst* 30:485–500
- Ashley J (1988) A linear bound for sliding block decoder window size. *Trans IEEE Inf Theory* 34:389–399
- Ashley J (1991) Resolving factor maps for shifts of finite type with equal entropy. *Ergod Theory Dyn Syst* 11:219–240
- Ashley J (1996) A linear bound for sliding block decoder window size, II. *Trans IEEE Inf Theory* 42:1913–1924
- Ashley J, Jaquette G, Marcus B, Seger P (1999) Runlength limited encoding/decoding with robust resync. US patent 5,969,649
- Aubrun N, Sablik M (2013) Simulation of effective subshifts by two-dimensional subshifts of finite type. *Acta Appl Math* 126:3563
- Barbero A, Rosnes E, Yang G, Ytrehus O (2014) Near-field passive RFID communication: channel model and code design. *IEEE Trans Commun* 62:1716–1726
- Béal M-P (1990) The method of poles: a coding method for constrained channels. *Trans IEEE Inf Theory* 36:763–772

- Béal M-P (1993) *Codage symbolique*. Masson, Paris
- Béal M-P (2003) Extensions of the method of poles for code construction. *Trans IEEE Inf Theory* 49:1516–1523
- Bedford T (1986) Generating special Markov partitions for hyperbolic toral automorphisms using fractals. *Ergod Theory Dyn Syst* 6:325–333
- Berger R (1966) *The undecidability of the domino problem*. *Memoirs of the American Mathematical Society*. AMS, Providence
- Berstel J, Perrin D (1985) *Theory of codes*. Academic, New York
- Blanchard F, Maass A, Nogueira A (2000) In: Blanchard F, Maass A, Nogueira A (eds) *Topics in symbolic dynamics and applications*. LMS lecture notes, vol 279. Cambridge University Press, Cambridge
- Blanchard P, Devaney R, Keen L (2004) *Complex dynamics and symbolic dynamics*. In: Williams S (ed) *Symbolic dynamics and its applications*. *Proceedings of symposia in applied mathematics*. AMS, Providence, pp 37–59
- Bowen R (1970) Markov partitions for axiom A diffeomorphisms. *Am J Math* 92:725–747
- Bowen R (1973) Symbolic dynamics for hyperbolic flows. *Am J Math* 95:429–460
- Bowen L (2012) Sofic entropy and amenable groups. *Ergod Theory Dyn Syst* 32:427–466
- Bowen L (2018) A brief introduction to sofic entropy theory, arXiv: 1711.02062v2
- Bowen R, Franks J (1977) Homology for zero-dimensional basic sets. *Ann Math* 106:73–92
- Bowen R, Lanford OE (1970) Zeta functions of restrictions of the shift transformation. *Proceedings of symposia in pure mathematics*, vol 14. AMS, Providence, pp 43–50
- Boyle M (1983) Lower entropy factors of sofic systems. *Ergod Theory Dyn Syst* 3:541–557
- Boyle M (1993) Symbolic dynamics and matrices. In: Brualdi R et al (eds) *Combinatorial and graph theoretic problems in linear algebra*. IMA volumes in mathematics and its applications, vol 50. Springer New York, New York, pp 1–38
- Boyle M (2007) Open problems in symbolic dynamics. *Contemporary Math*; for solutions and updates, see <http://www.math.umd.edu/mmb>
- Boyle M, Handelman D (1991) The spectra of nonnegative matrices via symbolic dynamics. *Ann Math* 133:249–316
- Boyle M, Krieger W (1986) Almost Markov and shift equivalent sofic systems. In: Alexander J (ed) *Proceedings of Maryland special year in dynamics 1986–87*. *Lecture notes in mathematics*, vol 1342. Springer, Berlin, pp 33–93
- Boyle M, Schmieding S (2016) Strong shift equivalence and the generalized spectral conjecture for nonnegative matrices. *Linear Algebra Appl* 498:231–243
- Boyle M, Schmieding S (2019) Strong shift equivalence and algebraic K-theory. *J Reine Angew Math* 752:63–104
- Boyle M, Marcus BH, Trow P (1987) Resolving maps and the dimension group for shifts of finite type. *Memoirs of the American Mathematical Society*, vol 377. AMS, Providence
- Boyle M, Buzzi J, Gomez-Aiza R (2006) Almost isomorphism for countable state Markov shifts. *J Reine Angew Math* 592:2347
- Boyle M, Pavlov R, Schraudner M (2010) Multi-dimensional sofic shifts without separation and their factors. *Trans Amer Math Soc* 362:4617–4653
- Boyle M, Carlsen TM, Eilers S (2018) Flow equivalence of sofic shifts. *Isr J Math* 225:111–146
- Burton R, Steif J (1994) Nonuniqueness of measures of maximal entropy for subshifts of finite type. *Ergod Theory Dyn Syst* 14(2):213–235
- Calkin N, Wilf H (1998) The number of independent sets in a grid graph. *SIAM J Discret Math* 11:54–60
- Ceccherini-Silberstein, T. and Coornaert, M. *Cellular automata and groups* (2010), Springer Berlin
- Cideciyan R, Evangelos E, Marcus B, Modha M (2001) Maximum transition run codes for generalized partial response channels. *IEEE J Sel Area Commun* 19:619–634
- Coven E, Paul M (1974) Endomorphisms of irreducible shifts of finite type. *Math Syst Theory* 8:167–175
- Coven E, Paul M (1975) Sofic systems. *Isr J Math* 20:165–177
- Coven E, Paul M (1977) Finite procedures for sofic systems. *Monats Math* 83:265–278
- Cover T, Thomas J (1991) *Elements of information theory*. Wiley, New York
- Desai A (2006) Subsystem entropy for $\mathbb{Z}^d$ sofic shifts. *Indag Math* 17:353–360
- Desai A (2009) A class of $\mathbb{Z}^d$ -subshifts which factor onto lower entropy full shifts. *Proc Am Math Soc* 137:2613–2621
- Devaney R (1987) *An introduction to chaotic dynamical systems*. Addison-Wesley, Reading
- Elishco O, Meyerovitch T, Schwartz M (2016) Semiconstrained systems. *IEEE Trans Inf Theory* 62:1688–1702
- Fischer R (1975a) Sofic systems and graphs. *Monats Math* 80:179–186
- Fischer R (1975b) Graphs and symbolic dynamics. *Colloq Math Soc János Bolyai: Top Inf Theory* 16:229–243
- Franaszek PA (1968) Sequence-state coding for digital transmission. *Bell Syst Tech J* 47:143–155
- Franaszek PA (1982) Construction of bounded delay codes for discrete noiseless channels. *J IBM Res Dev* 26:506–514
- Franaszek PA (1989) Coding for constrained channels: a comparison of two approaches. *J IBM Res Dev* 33:602–607
- Franks J (1984) Flow equivalence of subshifts of finite type. *Ergod Theory Dyn Syst* 4:5366
- Friedman J (1990) On the road coloring problem. *Proc Am Math Soc* 110:1133–1135
- Gomez R (2003) Positive K-theory for finitary isomorphisms of Markov chains. *Ergod Theory Dyn Syst* 23:1485–1504
- Hadamard J (1898) Les surfaces a courbures opposées et leurs lignes geodesiques. *J Math Pure Appl* 4:27–73

- Hasselblatt B, Katok A (1995) Introduction to the modern theory of dynamical systems. Cambridge University Press, Cambridge
- Hedlund GA (1939) The dynamics of geodesic flows. *Bull Am Math Soc* 45:241–260
- Hedlund GA (1944) Sturmian minimal sets. *Am J Math* 66:605–620
- Hedlund GA (1969) Endomorphisms and automorphisms of the shift dynamical system. *Math Syst Theory* 3:320–375
- Hochman M (2009) On the dynamics and recursive properties of multidimensional symbolic systems. *Invent Math* 176:131–167
- Hochman M, Meyerovitch T (2007) A characterization of the entropies of multidimensional shifts of finite type. *Ann Math* 171(2010):2011–2038
- Hollmann HDL (1995) On the construction of bounded-delay encodable codes for constrained systems. *Trans IEEE Inf Theory* 41:1354–1378
- Immink KAS (1994) EFMPPlus, 8–16 modulation code. US patent 5,696,505
- Immink KAS (2004a) Codes for mass data storage, 2nd edn. Shannon Foundation Press, Eindhoven
- Immink KAS (2004b) Codes for mass data storage systems, 2nd edn. Shannon Foundation Publishers, Eindhoven
- Jeandel E (2016) Computability in symbolic dynamics. *CiE, Paris*, pp 124–131
- Johnson A, Madden K (2005) Factoring higher-dimensional shifts of finite type onto full shifts. *Ergod Theory Dyn Syst* 25:811–822
- Karabed R, Siegel P, Soljanin E (1999) Constrained coding for binary channels with high intersymbol interference. *Trans IEEE Inf Theory* 45:1777–1797
- Kari J (2001) Synchronizing finite automata on Eulerian digraphs. *Springer Lect Notes Comput Sci* 2136:432–438
- Kastelyn PW (1961) The statistics of dimers on a lattice. *Physica A* 27:1209–1225
- Katok A (2007) Fifty years of entropy in dynamics 1958–2007. *J Modern Dyn* 4:545–596
- Keller G (1998) Equilibrium states in ergodic theory. Cambridge University Press, Cambridge
- Kenyon R (1996) The construction of self-similar tilings. *Geom Funct Anal* 6:471–488
- Kenyon R (2008) Lectures on dimers. <http://www.math.brown.edu/~rkenyon/papers/dimerlecturenotes.pdf>
- Kim KH, Roush FW (1979) Some results on decidability of shift equivalence. *J Comb Inf Syst Sci* 4:123–146
- Kim KH, Roush FW (1988) Decidability of shift equivalence. In: Alexander J (ed) *Proceedings of Maryland special year in dynamics 1986–87*. Lecture notes in mathematics, vol 1342. Springer, Berlin, pp 374–424
- Kim KH, Roush FW (1990) An algorithm for sofic shift equivalence. *Ergod Theory Dyn Syst* 10:381–393
- Kim KH, Roush FW (1999) Williams conjecture is false for irreducible subshifts. *Ann Math* 149:545–558
- Kim KH, Ormes N, Roush F (2000) The spectra of non-negative integer matrices via formal power series. *Am J Math Soc* 13:773–806
- Kitchens B (1998) Symbolic dynamics: one-sided, two-sided and countable state Markov chains. Springer, Berlin
- Kitchens B, Schmidt K (1988) Periodic points, decidability and Markov subgroups. In: Alexander JC (ed) *Dynamical systems: proceedings of the special year*. Springer lecture notes in mathematics, vol 1342. Springer Berlin Heidelberg, Berlin/Heidelberg, pp 440–454
- Kitchens B, Marcus B, Trow P (1991) Eventual factor maps and compositions of closing maps. *Ergod Theory Dyn Syst* 11:85–113
- Krieger W (1980a) On a dimension for a class of homeomorphism groups. *Math Ann* 252:87–95
- Krieger W (1980b) On dimension functions and topological Markov chains. *Invent Math* 56:239–250
- Krieger W (1982) On the subsystems of topological Markov chains. *Ergod Theory Dyn Syst* 2:195–202
- Krieger W (1983) On the finitary isomorphisms of Markov shifts that have finite expected coding time. *Wahrsch Z* 65:323–328
- Krieger W (1984) On sofic systems I. *Isr J Math* 48:305–330
- Krieger W (2018) On images of sofic systems, [arXiv:1101.1750](https://arxiv.org/abs/1101.1750)
- Lightwood S (2003/04) Morphisms form non-periodic Z^2 subshifts I and II. *Ergod Theory Dyn Syst* 23:587–609, 24:1227–1260
- Lind D (1984) The entropies of topological Markov shifts and a related class of algebraic integers. *Ergod Theory Dyn Syst* 4:283–300
- Lind D (1989) Perturbations of shifts of finite type. *SIAM J Discret Math* 2:350–365
- Lind D (1996) A zeta function for Z^d -actions. In: Pollicott M, Schmidt K (eds) *Proceedings of Warwick symposium on Z^d -actions*. LMS lecture notes, vol 228. Cambridge University Press, Cambridge, pp 433–450
- Lind D (2004) Multi-dimensional symbolic dynamics. In: Williams S (ed) *Symbolic dynamics and its applications*. Proceedings of symposia in applied mathematics, vol 60. AMS, Providence, pp 81–120
- Lind D, Marcus B (1995) An introduction to symbolic dynamics and coding. Cambridge University Press, Cambridge
- Lind D, Schmidt K (2002) Symbolic and algebraic dynamical systems. In: Hasselblatt B, Katok A (eds) *Handbook of dynamics systems*. Elsevier, Amsterdam, pp 765–812
- Manning A (1971) Axiom A diffeomorphisms have rational zeta functions. *Bull Lond Math Soc* 3:215–220
- Marcus B (1979) Factors and extensions of full shifts. *Monats Math* 88:239–247
- Marcus BH, Roth RM (1991) Bounds on the number of states in encoder graphs for input-constrained channels. *Trans IEEE Inf Theory* 37:742–758

- Marcus B, Tuncel S (1990) Entropy at a weight-per-symbol and embeddings of Markov chains. *Invent Math* 102:235–266
- Marcus B, Tuncel S (1991) The weight-per-symbol polytope and scaffolds of invariants associated with Markov chains. *Ergod Theory Dyn Syst* 11:129–180
- Marcus B, Tuncel S (1993) Matrices of polynomials, positivity, and finite equivalence of Markov chains. *J Am Math Soc* 6:131–147
- Marcus BH, Roth RM, Siegel PH (1998) Constrained systems and coding for recording chapter. In: Brualdi R, Huffman C, Pless V (eds) *Handbook on coding theory*. Elsevier, New York; updated version at <http://www.math.ubc.ca/~marcus/Handbook/>
- Markley N, Paul M (1981a) Matrix subshifts for $\mathbf{Z}^v$ symbolic dynamics. *Proc Lond Math Soc* 43:251–272
- Markley N, Paul M (1981b) Maximal measures and entropy for $\mathbf{Z}^v$ subshifts of finite type. In: Devaney R, Nitecki Z (eds) *Classical mechanics and dynamical systems*. Dekker notes, vol 70. Dekker, New York, pp 135–157
- Meester R, Steif J (2001) Higher-dimensional subshifts of finite type, factor maps and measures of maximal entropy. *Pac Math J* 200:497–510
- Morse M (1921) Recurrent geodesics on a surface of negative curvature. *Trans Am Math Soc* 22:84–100
- Morse M, Hedlund GA (1938) Symbolic dynamics. *Am J Math* 60:815–866
- Morse M, Hedlund GA (1940) Symbolic dynamics II, Sturmian trajectories. *Am J Math* 62:1–42
- Mouat R, Tuncel S (2002) Constructing finitary isomorphisms with finite expected coding time. *Isr J Math* 132:359–372
- Mozes S (1989) Tilings, substitutions and the dynamical systems generated by them. *J Anal Math* 53:139–186
- Mozes S (1992) A zero entropy, mixing of all orders tiling system. In: Walters P (ed) *Symbolic dynamics and its applications*. Contemporary mathematics, vol 135. AMS, Providence, pp 319–326
- Nagy Z, Zeger K (2000) Capacity bounds for the three-dimensional (0, 1) run length limited channel. *Trans IEEE Inf Theory* 46:1030–1033
- Nasu M (1986) Topological conjugacy for sofic systems. *Ergod Theory Dyn Syst* 6:265–280
- Ornstein D (1970) Bernoulli shifts with the same entropy are isomorphic. *Adv Math* 4:337–352
- Parry W (1964) Intrinsic Markov chains. *Trans Am Math Soc* 112:55–66
- Parry W (1977) A finitary classification of topological Markov chains and sofic systems. *Bull Lond Math Soc* 9:86–92
- Parry W (1979) Finitary isomorphisms with finite expected code-lengths. *Bull Lond Math Soc* 11:170–176
- Parry W (1991) Notes on coding problems for finite state processes. *Bull Lond Math Soc* 23:1–33
- Parry W, Schmidt K (1984) Natural coefficients and invariants for Markov shifts. *Invent Math* 76:15–32
- Parry W, Tuncel S (1981) On the classification of Markov chains by finite equivalence. *Ergod Theory Dyn Syst* 1:303–335
- Parry W, Tuncel S (1982) Classification problems in ergodic theory. LMS lecture notes, vol 67. Cambridge University Press, Cambridge
- Pavlov R (2011) Perturbations of multidimensional shifts of finite type. *Ergod Theory Dyn Syst* 31:483–526
- Petersen K (1989) *Ergodic theory*. Cambridge University Press, Cambridge
- Quas A, Sahin A (2003) Entropy gaps and locally maximal entropy in $\mathbf{Z}^d$ -subshifts. *Ergod Theory Dyn Syst* 23:1227–1245
- Quas A, Trow P (2000) Subshifts of multidimensional shifts of finite type. *Ergod Theory Dyn Syst* 20:859–874
- Radin C (1996) Miles of tiles. In: Pollicott M, Schmidt K (eds) *Ergodic theory of $\mathbf{Z}^d$ -actions*. LMS lecture notes, vol 228. Cambridge University Press, Cambridge, pp 237–258
- Robinson RM (1971) Undecidability and nonperiodicity for tilings of the plane. *Invent Math* 12:177–209
- Robinson EA (2004) Symbolic dynamics and tilings of $\mathbf{R}^d$. In: Williams S (ed) *Symbolic dynamics and its applications*. Proceedings of symposia in applied mathematics, vol 60. AMS, Providence, pp 81–120
- Roth R, Siegel P (2018) On parity-preserving constrained coding. In: *Proceedings of international symposium on information theory*, pp 1804–1808
- Rudolph D (1990) *Fundamentals of measurable dynamics*. Oxford University Press, Oxford
- Ruelle D (2004) *Thermodynamic formalism*, 2nd edn. Cambridge University Press, Cambridge
- Sarig O (2009) Lecture notes on thermodynamic formalism for topological Markov shifts. <http://www.weizmann.ac.il/math/sarigo/sites/math.sarigo/files/uploads/tdfnotes.pdf>
- Sarig O (2013) Symbolic dynamics for surface diffeomorphisms with positive entropy. *J Am Math Soc* 26:341–426
- Schmidt K (1984) Invariants for finitary isomorphisms with finite expected code lengths. *Invent Math* 76:33–40
- Schmidt K (1990) Algebraic ideas in ergodic theory. AMS-CBMS reg conference, vol 76. AMS, Providence
- Schmidt K (1995) *Dynamical systems of algebraic origin*. Birkhauser, Basel
- Schmidt K (2012) *Dynamical systems of algebraic origin*. Birkhauser (reprint of 1995 edition)
- Schwartz M, Bruck S (2008) Constrained codeds as networks of relations. *IEEE Trans Inf Theory* 54:2179–2195
- Seneta E (1980) *Non-negative matrices and Markov chains*, 2nd edn. Springer, Berlin
- Shannon C (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423, 623–656
- Sinai YG (1968) Markov partitions and C-diffeomorphisms. *Funct Anal Appl* 2:64–89
- Smale S (1967) Differentiable dynamical systems. *Bull Am Math Soc* 73:747–817
- Thomsen K (2004) On the structure of a sofic shift space. *Am Math Soc Trans* 356:3557–3619

- Thurston WP (1990) Groups, tilings, and finite state automata. Lecture notes, AMS colloquium lectures, American Mathematical Society
- Thurston WP (2014) Entropy in dimension one. *Frontiers in complex dynamics*, Princeton mathematical series, vol 51. Princeton University Press, Princeton
- Trachtman A (2007) The road coloring problem. *Israel J Math* 172:51–60
- Tuncel S (1981) Conditional pressure and coding. *Isr J Math* 39:101–112
- Tuncel S (1983) A dimension, dimension modules and Markov chains. *Proc Lond Math Soc* 46:100–116
- Wagoner J (1992) Classification of subshifts of finite type revisited. In: Walters P (ed) *Symbolic dynamics and its applications*. Contemporary mathematics, vol 135. AMS, Providence, pp 423–444
- Wagoner J (2004) Strong shift equivalence theory. In: Walters P (ed) *Symbolic dynamics and its applications*. Proceedings of symposia in applied mathematics, vol 60. AMS, Providence, pp 121–154
- Walters P (1982) *An introduction to ergodic theory*. Springer graduate texts in mathematics, vol 79. Springer, Berlin
- Walters P (1992) In: Walters P (ed) *Symbolic dynamics and its applications*. Contemporary mathematics, vol 135. AMS, Providence
- Ward T (1994) Automorphisms of Z^d -subshifts of finite type. *Indag Math* 5:495–504
- Weiss B (1973) Subshifts of finite type and sofic systems. *Monats Math* 77:462–474
- Williams RF (1973/74) Classification of subshifts of finite type. *Ann Math* 98:120–153; Erratum: *Ann Math* 99:380–381
- Williams S (2004a) Introduction to symbolic dynamics. In: Williams S (ed) *Symbolic dynamics and its applications*. Proceedings of symposia in applied mathematics, vol 60. AMS, Providence, pp 1–12
- Williams S (2004b) In: Williams S (ed) *Symbolic dynamics and its applications*. Proceedings of symposia in applied mathematics, vol 60. AMS, Providence

Operator Ergodic Theory

Guy Cohen¹ and Michael Lin²

¹School of Electrical Engineering, Ben-Gurion University, Beer-Sheva, Israel

²Department of Mathematics, Ben-Gurion University, Beer-Sheva, Israel

Article Outline

Glossary

Definition of the Subject

Introduction

The Mean Ergodic Theorem

Rates of Convergence

Uniform Ergodic Theorems

Strong Cesàro Convergence

Weak Stability and Mixing

Stability

Averaging Along Subsequences

General Averaging Methods

Modulated Ergodic Theorems

Resolvent Conditions and Growth of Powers

Continuous Time (C_0 -semigroups)

Bibliography

Glossary

Blum-Hanson property Norm convergence to zero of all averages $\frac{1}{n} \sum_{j=1}^n T^{k_j} x$ along any increasing $\{k_j\} \subset \mathbb{N}$ whenever $T^n x \rightarrow 0$ weakly.

Cesàro bounded operator An operator T with $\sup_n \left\| \frac{1}{n} \sum_{k=0}^{n-1} T^k \right\| < \infty$.

Coboundary (of an operator) A vector $y \in (I - T)X$, i.e., such that $y = x - Tx$ for some $x \in X$.

Contraction An operator T with $\|T\| \leq 1$.

C_0 -semigroup/one-parameter semigroup A family of bounded operators $\{T(t)\}_{t \geq 0}$ defined on X such that $T(t+s) = T(t)T(s)$ for $t, s \geq 0$, $T(0) = I$, and $T(t)x$ is continuous on $[0, \infty)$ for every $x \in X$.

Ergodic measure-preserving transformation A mpt such that the only sets A with $\theta^{-1}A = A$ a.e are null sets or complements of null sets.

Infinitesimal generator of a C_0 -semigroup The operator $Ax := \lim_{t \rightarrow 0+} t^{-1}(T(t)x - x)$ with domain $\mathcal{D}(A) := \{x \in X : Ax \text{ exists}\}$.

Koopman operator The operator on $L_p(\Omega, \Sigma, \mu)$, induced by a measure-preserving transformation θ , defined by $Tf = f \circ \theta$.

Markov-Feller operator A Markov operator on bounded measurable functions of a compact Hausdorff space which preserves continuity.

Markov operator The operator induced on bounded measurable functions on (Ω, Σ) by a transition probability $P(t, A)$, defined by $Pf(t) = \int f(s)P(t, ds)$.

Measure (probability)-preserving transformation (mpt) A measurable mapping θ of a measure (probability) space (Ω, Σ, μ) satisfying $\mu(\theta^{-1}A) = \mu(A)$ for any $A \in \Sigma$.

Minimal topological dynamical system A topological dynamical system (K, τ) such that for every $s \in K$ the orbit $\{\tau^n s\}$ is dense.

Positive operator (on a Banach lattice) An operator which maps the positive cone of a Banach lattice into itself.

Power-bounded operator An operator T with $\sup_n \|T^n\| < \infty$.

Operator A bounded linear operator on a (real or complex) Banach space X .

Operator semigroup A family S of operators such that $TS \in S$ for $T, S \in S$.

Orbit of a vector under an operator semigroup S The set $\{Tx : T \in S\}$.

(Semi)flow A family of transformations $\{\theta_t\}_{t \in \mathbb{R}}$ ($\{\theta_t\}_{t \geq 0}$) such that $\theta_{t+s} = \theta_t \circ \theta_s$, with θ_0 the identity map.

Stolz region The closed convex hull (in $\mathbb{C}$) of the point 1 and a disk of radius $r < 1$.

Transition probability A function $P(t, A)$ on $\Omega \times \Sigma$ (where (Ω, Σ) is a measurable space), such that $P(t, \cdot)$ is a probability on Σ for $t \in \Omega$ and $P(\cdot, A)$ is measurable for every $A \in \Sigma$.

Transformation A mapping of a set into itself.

Topological dynamical system A compact Hausdorff space K with a continuous map τ of K to itself.

Uniform ergodicity Convergence of the averages $\frac{1}{n} \sum_{k=0}^{n-1} T^k$ in the operator norm (uniformly on the unit ball of the Banach space).

Uniquely ergodic topological dynamical system A topological dynamical system (K, τ) with a unique probability measure μ satisfying $\mu \circ \tau^{-1} = \mu$.

(Weak) mean ergodic theorem (weak) convergence of the averages $\frac{1}{n} \sum_{k=0}^{n-1} T^k x$ for every x in the Banach space X on which the operator T is defined.

(Weak) stability (Weak) convergence of $T^n x$ for every $x \in X$.

(Weakly) almost periodic operator An operator T such that for every $x \in X$ the orbit $\{T^n x\}$ is (weakly) conditionally compact.

(Weakly) quasi-compact operator An operator T such that for some $n > 0$ there exists a (weakly) compact operator Q satisfying $\|T^n - Q\| < 1$.

Definition of the Subject

Operator ergodic theory deals with the asymptotic behavior of the powers of a power-bounded operator, or a bounded one-parameter semigroup of operators, defined on a (real or complex) Banach space. We will focus on the theory for powers of a single operator, and with the growth of the norms of certain operators which are only Cesàro bounded. The asymptotic behavior is studied in different operator topologies and in various modes of convergence. However, almost everywhere convergence of positive operators on L_p spaces is not discussed methodically in this chapter, since this is done in other chapters of this volume.

The topic of ergodic theorems emerged from the mean ergodic theorem for unitary operators in complex Hilbert spaces of von Neumann (1932) and the pointwise ergodic theorem of G.D. Birkhoff (1931) for measure-preserving transformations. We quote from the survey of

Halmos (1949a): “It was very quickly recognized that the proper general framework for von Neumann’s mean ergodic theorem lay in the direction of Hilbert spaces and Banach spaces, whereas the extent of generality suitable to Birkhoff’s theorem was to be found in the concept of a measure space.” Thus, the origin of operator ergodic theory is in the mean ergodic theorem, although the uniform distribution theorem of Weyl (1916) can be restated as an ergodic theorem. The ergodic theory of measure-preserving transformations and pointwise limit theorems for some classes of positive operators are discussed in other chapters of this volume.

Birkhoff and von Neumann were both motivated by a problem in statistical mechanics, Boltzmann’s “ergodic hypothesis,” so accordingly their theorems were formulated and proved in continuous time, i.e., for a flow of invertible measure-preserving transformations rather than for the iterates of one such transformation (discrete time). See Petersen (1996) for comparison of the significance in scientific applications of these two theorems.

Operator ergodic theory has applications in the study of measure-preserving transformations, topological dynamics, and Markov operators, but it is also inspired by results first obtained in these areas. Results for C_0 -semigroups (continuous time) have applications also in partial differential equations.

Introduction

Let $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$ be the unit circle in the complex plane, and let $\alpha \in [0, 1)$ be *irrational*. The transformation $\theta z := e^{2\pi i \alpha} z, z \in \mathbb{T}$ corresponds to a rotation of the circle by the angle $2\pi\alpha$. In 1884, Kronecker proved that for each z the orbit $\{\theta^n z\}$ is dense in $\mathbb{T}$. Weyl (1910) extended it in his famous uniform distribution theorem, which, using Weyl (1916), can be stated as a mean ergodic theorem in $C(\mathbb{T})$.

Theorem 1 Let $\alpha \in [0, 1)$ be irrational, and for f continuous on $\mathbb{T}$ define $Tf(z) = f(e^{2\pi i \alpha} z)$. Then

$\frac{1}{n} \sum_{k=1}^n T^k f(z)$ converges uniformly to the constant $\int f d\mu$, where μ is the normalized Lebesgue measure on $\mathbb{T}$.

The result is immediate for trigonometric polynomials, which by Fejér's theorem are dense in $C(\mathbb{T})$; this proves the theorem. Moreover, trigonometric polynomials are dense in $L_2(\mathbb{T}, \mu)$, and we obtain the following.

Corollary 2 Let $\alpha \in [0, 1)$ be irrational, and let $Tf(z) = f(e^{2\pi i \alpha} z)$ be the Koopman operator on $L_2(\mathbb{T}, \mu)$. Then for every $f \in L_2(\mathbb{T}, \mu)$ we have $\left\| \frac{1}{n} \sum_{k=1}^n T^k f - \int f d\mu \right\|_2 \rightarrow 0$.

Now let θ be any invertible measure-preserving transformation of a probability space (Ω, Σ, μ) . Then the Koopman operator $Tf = f \circ \theta$ is unitary on $L_2(\Omega, \Sigma, \mu)$, an observation which led von Neumann (1932) to use the spectral theorem to prove the following mean ergodic theorem.

Theorem 3 Let T be a unitary operator on a complex Hilbert space $\mathcal{H}$, and let E be the orthogonal projection on the space $\mathcal{F}(T)$ of T -fixed vectors. Then for any $x \in \mathcal{H}$ we have $\left\| \frac{1}{n} \sum_{k=1}^n T^k x - Ex \right\| \rightarrow 0$.

Koopman and von Neumann (1932) proved that if T in Theorem 3 has no eigenvalues except 1, then

$$Ex = 0 \Rightarrow \frac{1}{n} \sum_{k=0}^{n-1} |\langle T^k x, y \rangle| \rightarrow 0 \text{ for every } y \in \mathcal{H}, \quad (1)$$

but $T^n x$ need not converge weakly to 0. Hopf (1932) characterized (1) when T is the Koopman operator of an ergodic probability-preserving transformation τ , by ergodicity of $(\tau \times \tau)(u, v) := (\tau u, \tau v)$ on $(\Omega \times \Omega, \mu \times \mu)$; in Hopf (1937), he called such τ weak mixing, while τ was called mixing if $T^n x \rightarrow Ex$ weakly.

By a strange coincidence, Banach's book (1932) was published in the same year, and it was realized that von Neumann's theorem can be generalized "to more general Banach spaces and to more general classes of operators," quoting Kakutani (1950).

In Hopf (1937), F. Riesz proved elementarily that when T is a contraction of a Hilbert space, $Tx = x$ if and only if $T^*x = x$, and extended Theorem 3 to contractions. Visser (1938) proved weak mean ergodicity of power-bounded operators in Hilbert space. G. Birkhoff (1939) proved the mean ergodic theorem for contractions in uniformly convex spaces, extending the L_p result of F. Riesz (1938).

In their study of Markov chains, Kryloff and Bogoliouboff (1937) presented the following.

Theorem 4 Let $P(t, A)$ be a transition probability on (Ω, Σ) and T the Markov operator defined on the space of bounded measurable functions by $Tf(t) = \int f(s)P(t, ds)$. If there exist $n \geq 1$ and a compact operator Q , such that $\|T^n - Q\| < 1$, then T is uniformly ergodic – the averages $\frac{1}{n} \sum_{k=1}^n T^k$ converge in operator norm.

The Mean Ergodic Theorem

An operator T on a (real or complex) Banach space X is called (weakly) mean ergodic if for every $x \in X$ the averages $M_n x := \frac{1}{n} \sum_{k=0}^{n-1} T^k x$ converge (weakly); the limit operator E is a projection on the subspace $\mathcal{F}(T) := \{y \in X : Ty = y\}$ of fixed points of T . A (weakly) mean ergodic operator must be Cesàro bounded ($\sup_n \|M_n\| < \infty$) and, since $T^n x = (n+1)M_{n+1}x - nM_n x$, must satisfy $\frac{1}{n} T^n x \rightarrow 0$ (weakly) for every x .

Lemma 5 If $\frac{1}{n} T^n x \rightarrow 0$ weakly and for some subsequence $\{n_j\}$ we have $M_{n_j} x \rightarrow y$ weakly, then $Ty = y$.

Proposition 6 Let T be Cesàro bounded on X satisfying $\left\| \frac{1}{n} T^n x \right\| \rightarrow 0$ for every $x \in X$. Then the following are equivalent for $x \in X$:

- (i) $\|M_n x\| \rightarrow 0$.
- (ii) $M_{n_j} x \rightarrow 0$ weakly for some increasing subsequence $\{n_j\}$.
- (iii) $\phi(x) = 0$ for every functional $\phi \in \mathcal{F}(T^*) \subset X^*$.
- (iv) $x \in \overline{(I - T)X}$.

The proof of (iii) implies (iv) uses the Hahn-Banach theorem. (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) are obvious, and (iv) $\Rightarrow$ (i) is easy. Proposition 6 is part of the proof in Yosida (1938) of the following mean ergodic theorem, proved independently in Kakutani (1938).

An operator T on X is called *weakly almost periodic* if for every $x \in X$ the orbit $\{T^n x\}_n$ is weakly conditionally compact.

Theorem 7 *Let T be a weakly almost periodic operator on X . Then T is mean ergodic.*

Since the Krein-Šmulian theorem appeared only in 1940, Yosida assumed that T is power-bounded and that $\{M_n x\}$ is weakly compact. Both assumptions follow from weak almost periodicity.

The operator in Theorem 1 is weakly almost periodic on $C(\mathbb{T})$. Weak almost periodicity is not necessary for mean ergodicity of a power-bounded operator; Sine (1976) has a mean ergodic positive contraction T on $C(K)$ such that T^2 is not mean ergodic, hence T is not weakly almost periodic. Several other examples are presented by Gerlach and Glück (2019). However, if T is a positive contraction on a Banach lattice with order continuous norm, mean ergodicity of T implies that of T^2 , by Derriennic and Krengel (1981). Komorník (1993) proved that a *positive* contraction on L_1 is weakly almost periodic if (and only if) it is mean ergodic.

Corollary 8 *Every power-bounded operator on a reflexive Banach space is mean ergodic.*

Corollary 8 was proved by Lorch (1939), independently of Kakutani and of Yosida. Kakutani and Yosida both saw that for proving Corollary 8 you really need only weak almost periodicity of the operator. The results of von Neumann, Riesz, Visser, and Birkhoff mentioned in the introduction are special cases of Corollary 8.

Eberlein (1949) extended the mean ergodic theorem to general operator semigroups $\mathcal{S}$, replacing the Cesàro averages by what he called *invariant integrals* – a family $\{M_\alpha\}$ of operators on the space X , indexed by α in a directed set, such that for any $x \in X$, $M_\alpha x$ is in the closed convex hull of the orbit $\{Tx : T \in \mathcal{S}\}$ of x , $\sup_\alpha \|M_\alpha\| < \infty$, and $\lim_\alpha M_\alpha(I - T)x = \lim_\alpha (I - T)M_\alpha x = 0$ for any T in

the semigroup and $x \in X$. Eberlein proved a general ergodic theorem on the convergence of $M_\alpha x$, from which he deduced Theorem 7.

The following was proved by Krotkov and Halperin (1953).

Proposition 9 *Let T be Cesàro bounded on a reflexive Banach X . If $\frac{1}{n}T^n x \rightarrow 0$ (weakly) for every $x \in X$, then T is (weakly) mean ergodic.*

Combining Lemma 5 and Proposition 6, we obtain that a power-bounded operator is mean ergodic if (and only if) it is weakly mean ergodic. However, this is false in general; Derriennic (2000) has an example of a weakly mean ergodic operator on a Hilbert space which is not mean ergodic (so $\limsup_n \frac{1}{n}\|T^n\| > 0$). A general method for obtaining examples like Derriennic's was given by Tomilov and Zemánek (2004).

Hille (1945) constructed, on $L_1[0, 1]$, the first example of a mean ergodic operator which is not power-bounded, with $\|T^n\| \sim O(n^{1/4})$. Kornfeld and Kosek (2003) constructed, for any $\gamma \in (0, 1)$, a *positive* mean ergodic operator on $L_1[0, 1]$ satisfying $\lim_n n^{\gamma-1}\|T^n\| = \infty$. Kosek (2011) constructed a mean ergodic operator on L_1 with $\limsup n^{-1}\|T^n\| > 0$. Cesàro boundeness alone does not imply mean ergodicity, even if X is finite-dimensional, by a simple example of Assani (1986). However, Émilion (1985) proved that any Cesàro-bounded *positive* operator on a reflexive Banach lattice is mean ergodic ($n^{-1}T^n \rightarrow 0$ strongly is not assumed, but follows).

Lemma 5 and Proposition 6 yield the following:

Theorem 10 *Let T be weakly mean ergodic on X . Then*

$$X = \mathcal{F}(T) \oplus \overline{(I - T)X}, \quad (2)$$

and the limit operator E is the projection on $\mathcal{F}(T)$ corresponding to (2).

Conversely, if T is Cesàro bounded and $\frac{1}{n}T^n x \rightarrow 0$ (weakly) for every $x \in X$ and (2) holds, then T is (weakly) mean ergodic.

For T as in Proposition 6, it is easy to show that the set of points $x \in X$ such that $(M_n x)$ converges is a closed subspace, which equals the right-hand side of (2).

Using the decomposition (2), we obtain an easy proof, due to Satō (1979), of the following theorem of Sine (1970), originally proved for contractions.

Theorem 11 *Let T be Cesàro bounded with $\frac{1}{n}T^n x \rightarrow 0$ for every $x \in X$. Then T is mean ergodic if and only if the fixed points of T separate the fixed points of T^* .*

Proposition 6 yields easily that for T as in Theorem 11, the fixed points of T^* always separate those of T , so if X is reflexive, we obtain Corollary 8 from Theorem 11.

Corollary 12 *Let (K, τ) be a uniquely ergodic topological dynamical system. Then the operator $Tf = f \circ \tau$ on $C(K)$ is mean ergodic.*

The corollary is due to Oxtoby (1952). Theorem 1 follows from Theorem 7 or from Corollary 12.

Theorem 13 *Let $0 \leq r_n \rightarrow 0$, and let T be mean ergodic, with $Ex = \lim M_n x$. If for every $x \in X$ there is c_x such that $\|M_n x - Ex\| \leq c_x r_n$, then $\|M_n - E\| \rightarrow 0$ (T is uniformly ergodic).*

Theorem 13 is a consequence of the uniform boundedness principle. When T acts on a complex Banach space, the decomposition (2) yields that when T is uniformly ergodic, 1 is an isolated point in the spectrum of T (see the section *Uniform ergodic theorems*). Since, by A. Ionescu Tulcea (1963) and Foias (1964), the spectrum of the L_2 Koopman operator T of an invertible ergodic probability-preserving transformation of a Lebesgue space is the whole unit circle $\mathbb{T}$, T has no uniform rate of convergence in the mean ergodic theorem. A constructive proof for this was given by Krengel (1978); see also Kakutani and Petersen (1981).

Motivated by Euler's formal approach to Fourier series, Wintner (1945) studied the existence of solutions $g \in L_1$ to the *coboundary equation* $f = g - g \circ \theta$, where θ is as in Theorem 1 and $f \in L_2$

(or L_1); Anosov (1973) proved that there exists $f \in C(\mathbb{T})$ such that $f = g - g \circ \theta$ with g non-integrable. Gottschalk and Hedlund (1955) proved that if (K, τ) is a minimal topological dynamical system and $Tf = f \circ \tau$ on $C(K)$, then $f \in (I - T)C(K)$ if (and only if) $\sup_n \|\sum_{k=0}^{n-1} T^k f\|_\infty < \infty$. Inspired by this result, Browder (1958) proved the following:

Theorem 14 *Let T be power-bounded on a reflexive Banach space X . Then*

$$y \in (I - T)X \quad \text{if and only if} \quad \sup_n \left\| \sum_{k=0}^{n-1} T^k y \right\| < \infty. \quad (3)$$

Put $S_n := \sum_{k=0}^{n-1} T^k$. The seemingly weaker condition $\sup_N \frac{1}{N} \sum_{n=1}^N \|S_n y\|^2 < \infty$ also implies $y \in (I - T)X$. This condition was shown in Robinson (1960) to imply $y \in (I - T)X$ when T is unitary; see also Kozma and Lev (2011). Cuny and Weber (2018) and Volný (2018) proved that if T is a contraction on a Hilbert space, then $\sum_{k=0}^\infty k \|T^k y\|^2 < \infty$ implies $y \in (I - T)X$. Iterative methods for solving Poisson's equation $(I - T)x = y$ were studied by Reich (1973).

Lin and Sine (1983) extended Browder's result to dual operators, proving (3) when $X = Y^*$ and $T = S^*$, with S power-bounded on Y ; see also Devys (2012). They also proved (3) for contractions on L_1 . See Aaronson and Weiss (2000) for related results. Kornfeld and Lin (1997) proved (3) for irreducible Markov-Feller operators on $C(K)$, thus extending the topological dynamics result of Gottschalk and Hedlund (1955). Lin and Suci (2015) studied the equation $(I - T)x = y$ when T is weakly mean ergodic.

Square functions were originally used in the study of a.e. convergence of orthogonal expansions; see the survey of Stein (1982). Motivated by a square function inequality by Jones, Ostrovskii, and Rosenblatt (1996) for contractions in Hilbert space, and by its extension by Jones et al. (1998) to certain isometries in L_p , Avigad and Rute (2015) obtained the following p -variation theorem in uniformly convex spaces.

Theorem 15 *Let X be isomorphic to a uniformly convex Banach space. Then there exists $p \geq 2$ such that for every power-bounded T satisfying $\inf_n \|T^n x\| \geq B \|x\|$ for some $B > 0$ and every $x \in X$, there exists C (depending on X, p , and T) such that for every increasing sequence $\{n_k\}$ we have*

$$\sum_{k=1}^{\infty} \|(M_{n_{k+1}} - M_{n_k})x\|^p \leq C \|x\|^p \quad \forall x \in X.$$

Kohlenbach and Leuştean (2009) used proof theory to obtain a quantitative proof of the mean ergodic theorem of G. Birkhoff (1939).

Let T be invertible on X reflexive with $\sup_{n \in \mathbb{Z}} \|T^n\| < \infty$. Then $\left\| \frac{1}{n} \sum_{k=\ell}^{\ell+n} T^k x - Ex \right\| \rightarrow 0$ for any $\ell \in \mathbb{Z}$. Motivated by Cotlar's (1955) a.e. convergence of the ergodic Hilbert transform of a probability-preserving transformation and by Petersen (1983), Campbell (1986) used the spectral theorem and the unitary dilation theorem to prove the following.

Theorem 16 *Let T be a contraction on a complex Hilbert space $\mathcal{H}$. Then for every $x \in \mathcal{H}$, the limit $Hx = \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k} (T^k x - T^{*k} x)$ exists, and H is a bounded linear operator on $\mathcal{H}$*

Berkson and Gillespie (1987) proved that if T is invertible on L_p , $1 < p < \infty$, with $\sup_{n \in \mathbb{Z}} \|T^n\| < \infty$, then $\sum_{k=1}^n \frac{1}{k} (T^k x - T^{-k} x)$ converges for every $x \in L_p$.

Sucheston (1976) asked whether a Banach space, in which every contraction (or every power-bounded operator) is mean ergodic, must be reflexive. Building upon a partial result of Zaharopol (1986), Emel'yanov (1997) proved that a Banach lattice is reflexive if (and only if) every power-bounded operator is mean ergodic. Fonf et al. (2001) proved that *if a Banach space has a basis and all power-bounded operators are mean ergodic, then the space must be reflexive*; additional equivalent conditions were obtained by Farkas and Kreidler (2021). Since a basis is assumed, this result does not include Emel'yanov's. An analogue for C_0 -semigroups was proved by Mugnolo (2004).

A nonreflexive Banach space with all contractions means ergodic was exhibited by Fonf et al. (2010).

For any power-bounded T we have (using Proposition 6 for the second inclusion)

$$(I - T)X \subset \left\{ x \in X : \sup_n \left\| \sum_{k=0}^{n-1} T^k x \right\| < \infty \right\} \subset \overline{(I - T)X}, \quad (4)$$

Theorem 14 yields equality in the first inclusion when X is reflexive. Fonf et al. (2011) proved that if X has a basis and for every power-bounded T there is equality in the first inclusion of (4), then X is reflexive.

Bermúdez et al. (2000) studied operators such that $T^n x/n \rightarrow 0$ implies that $M_n(T)x$ converges (a "local" mean ergodicity) and proved that if T^k has this property, so does T . Radjavi et al. (2003) studied local ergodic properties for T which is not Cesàro bounded. Their main result is:

Theorem 17 *Let T be a compact operator on a (real or complex) Banach space X , and let $x \in X$. (i) If $n^{-1} \|T^n x\| \rightarrow 0$ and $\sup_n \|M_n x\| < \infty$, then $M_n x$ converges in norm to a fixed point of T . (ii) If for some subsequence $\sup_k \|T^{n_k} x\| < \infty$, then $M_{n_k} x$ converges.*

Rates of Convergence

Let T be mean ergodic. When T is uniformly ergodic, (2) yields that S , the restriction of T to $Y := \overline{(I - T)X}$, satisfies $\|M_n(S)\| \rightarrow 0$. For large n , we have that $I - M_n(S)$ is invertible on Y , so $I - S$ is invertible, i.e., $I - T$ is invertible on $\overline{(I - T)X}$, so $(I - T)X$ is closed, and $\|M_n - E\| = O(\frac{1}{n})$. When T is not uniformly ergodic, by Theorem 13 there is no uniform rate of convergence, so we may find $x \in X$ with $\|M_n x - Ex\| \rightarrow 0$ arbitrarily slowly. However, Butzer and Westphal (1971) observed that $\|M_n x\| = o(\frac{1}{n})$ implies $x = 0$, so the fastest possible rate for $x \in \overline{(I - T)X}$ is $O(\frac{1}{n})$, when $x \in (I - T)X$; by Theorem 14, if X is reflexive, this rate is attained only when $x \in (I - T)X$.

For a unitary operator U on a complex Hilbert space $\mathcal{H}$, the spectral theorem entails a strong connection between the von Neumann ergodic theorem and the spectral measure of U . This connection was used to study conditions on σ_f , the spectral measure of $f \in \mathcal{H}$ with respect to U , which yield rates for the convergence of $M_n f$. In this context, we mention the works of Gaposhkin (1998) and of Kachurovskii and Podvigin (2016), which contain references to earlier works. We present a representative result from Kachurovskii and Sedalishchev (2010) and Kachurovskii and Podvigin (2016).

Theorem 18 *Let $f \in \mathcal{H}$ with spectral measure σ_f defined on $(-\pi, \pi]$, and assume $Ef = 0$. Then for any $a \in (0, \pi]$ and for any n , we have*

$$\|M_n f\|^2 \geq \left(\frac{\sin(a/2)}{a/2} \right)^2 \sigma_f \left(-\frac{a}{n}, \frac{a}{n} \right],$$

where the constant is best possible. Put $S_k = \sigma_f \left(-\frac{a}{k}, \frac{a}{k} \right]$ and $\Delta S_k = S_k - S_{k+1}$; then

$$\frac{4}{\pi^2} S_n \leq \|M_n f\|^2 \leq S_n + \frac{1}{n} \sum_{k=1}^{n-1} (k+1)^2 \Delta S_k.$$

Let $\alpha \in [0, 2)$. Then the rate $\|M_n f\|^2 \leq Bn^{-\alpha}$ holds if and only if for some $A > 0$ we have $\sigma_f(-\delta, \delta] \leq A\delta^\alpha$ for every $\delta \in (0, \pi]$.

Butzer and Westphal (1972) obtained for T power-bounded mean ergodic, but not uniformly ergodic, the following approximation estimate of $M_n y$ for $y \in \overline{(I - T)X}$: There exists $B > 0$ such that $\|M_n y\| \leq B \inf_{x \in (I - T)X} \left\{ \|y - x\| + \frac{1}{n+1} \|(I - T)^{-1}x\| \right\}$. Nasri-Roudsari et al. (1995) proved the optimality of this type of estimates.

Let T be power-bounded on a Banach space X (so T is a contraction in the equivalent norm $\|x\| := \sup_{n \geq 0} \|T^n x\|$). For $0 < \alpha < 1$, Derriennic and Lin (2001) used the power series $(1 - t)^\alpha = 1 - \sum_{j=1}^{\infty} a_j^{(\alpha)} t^j$, $0 \leq t \leq 1$, to define the fractional power $(I - T)^\alpha = I - \sum_{j=1}^{\infty} a_j^{(\alpha)} T^j$ (fractional powers of generators of continuous one-parameter semigroups had been defined differently more than 40 years earlier). They proved that

$\overline{(I - T)^2 X} = \overline{(I - T)X}$, but if T is not uniformly ergodic, the linear manifolds $(I - T)^\alpha X$ are all different, and proved their following characterization.

Theorem 19 *Let T be a mean ergodic contraction on a Banach space X and $0 < \alpha < 1$. Then $y \in (I - T)^\alpha X$ if and only if the series $\sum_{j=1}^{\infty} \frac{1}{j^{1-\alpha}} T^j y$ converges strongly. Thus $\|M_n y\| = o(1/n^\alpha)$ for $y \in (I - T)^\alpha X$.*

They concluded that $y \in (I - T)^\alpha X$ for some $\alpha \in (0, 1)$ implies convergence of the one-sided ergodic Hilbert transform $Hy := \sum_{j=1}^{\infty} \frac{T^j y}{j}$. Cuny (2009) proved for positive contractions of a fixed L_p , $p > 1$, that norm convergence of Hf implies its almost everywhere convergence. Additional properties of H and its domain of definition were given by Cohen et al. (2010) and by Haase and Tomilov (2010). Using a functional calculus, Gomilko et al. (2011) proved:

Theorem 20 *If T is power-bounded and the series $\sum_{j=1}^{\infty} \frac{T^j y}{j}$ converges weakly, then $\|M_n y\| = o(1/\log n)$; this rate is optimal.*

Halmos (1949b) proved that if T is the Koopman operator of an invertible ergodic measure preserving transformation of a nonatomic probability space, there exists $f \in \overline{(I - T)L_2}$ for which $\sum_{j=1}^{\infty} \frac{T^j f}{j}$ does not converge in L_2 -norm. Cuny (2009) proved that if that series converges in norm (even without invertibility of the transformation), then it converges a.e.

Let T be invertible on L^p , $1 < p < \infty$, with $\sup_{n \in \mathbb{Z}} \|T^n\| < \infty$; Cuny (2010) used the spectral integration of Berkson and Gillespie (1987) to obtain some conditions for the L_p convergence of $\sum_{j=1}^{\infty} \frac{1}{j^{1-\alpha}} T^j y$, $0 \leq \alpha < 1$.

Uniform Ergodic Theorems

The convergence of the averages of a mean ergodic operator need not be in operator norm, even in reflexive spaces; this is the case of the L_2 Koopman operator of an ergodic invertible probability-preserving transformation on $[0, 1]$

with Lebesgue measure. When the averages $M_n(T)$ converge in operator norm, we say that T is *uniformly ergodic*.

Fonf et al. (2001) proved that if X is a Banach space with basis such that every power-bounded operator is uniformly ergodic, then X is finite-dimensional. Thus conditions for uniform ergodicity of an operator, with or without prior assumption of mean ergodicity, are of interest.

We have already noted that if T is uniformly ergodic, then $Y := (I - T)X$ is closed and $I - T$ is invertible on Y . When X is over $\mathbb{C}$, this yields that for λ near 1 the restriction of $\lambda I - T$ to Y is invertible. The decomposition (2) yields that those λ are in the resolvent $\rho(T)$; thus, if $1 \in \sigma(T)$, then it is isolated in the spectrum. See Theorem 22 below for additional information.

In their study of Markov chains, Kryloff and Bogoliouboff (1937) introduced a new concept, which applies also to general operators. An operator T on a Banach space is called *quasi-compact* if there exist $n \geq 1$ and a compact operator Q , such that $\|T^n - Q\| < 1$. Yosida and Kakutani (1938) proved that a *power-bounded weakly quasi-compact operator* (Q weakly compact) is *mean ergodic*. Yosida and Kakutani (1941) observed that T is (weakly) quasi-compact if and only if there exists a sequence of (weakly) compact operators Q_n such that $\|T^n - Q_n\| \rightarrow 0$; they showed that the condition of Doeblin (1937) on Markov operators implies quasi-compactness and proved the following theorem, of which Theorem 4 is a special case.

Theorem 21 *Let T be a power-bounded quasi-compact operator on a complex Banach space X . Then T is uniformly ergodic. More precisely, T has only finitely many eigenvalues of modulus one (if any), say λ_j , $j = 1, \dots, k$, each of finite multiplicity, and there exist projections T_j , $j = 1, \dots, k$ and a quasi-compact S with $\|S^n\| \leq \frac{M}{(1+\varepsilon)^n}$ for some $\varepsilon > 0$, such that*

$$T^n = \sum_{j=1}^k \lambda_j^n T_j + S^n \quad n = 1, 2, \dots$$

with $TT_j = T_jT = \lambda_j T_j$, $T_j T_i = 0$ for $i \neq j$, and $T_j S = S T_j = 0$.

When T is as in Theorem 21, it is implied in the proof that its peripheral spectrum $\sigma(T) \cap \mathbb{T}$ consists only of (finitely many) eigenvalues; this follows from Theorem 22 below, since λT is quasi-compact, hence uniformly ergodic, for every $\lambda \in \mathbb{T}$.

Brunel and Revuz (1974) observed that T is quasi-compact if and only if $T = S + Q$, with spectral radius $r(S) < 1$ and Q with finite-dimensional range.

It follows from the work of Eberlein (1949) that if T is *power-bounded and a convex power-series* $S := \sum_{k=1}^{\infty} a_k T^k$ ($a_k \geq 0$, $\sum_k a_k = 1$), is (weakly) *quasi-compact*, then T is *mean ergodic*, *uniformly ergodic* when S is *quasi-compact*; this extends the results of Yosida and Kakutani (1938, 1941).

When T is uniformly ergodic, $(I - T)X$ is closed, so when T is power-bounded we have $\|M_n - E\| = O(1/n)$. Yoshimoto (1993) studied the uniform rate of convergence when T is quasi-compact but not power-bounded, and usual averages are replaced by $C(\alpha)$ averages; in Yoshimoto (1996), he studied the power-bounded case.

As an application of his operational calculus, Dunford (1943a) obtained the following.

Theorem 22 *Let T on a complex Banach space X satisfy $n^{-1}\|T^n\| \rightarrow 0$. Then the following are equivalent:*

- (i) T is uniformly ergodic.
- (ii) Either $1 \in \rho(T)$, or 1 is a simple pole of the resolvent $R(\lambda, T) := (\lambda I - T)^{-1}$.
- (iii) $(I - T)^2 X$ is closed.

Note that even for T power-bounded, 1 isolated in $\sigma(T)$ does not imply that 1 is a pole, e.g., $T = I - V$ on $L_2[0, 1]$ (V the Volterra operator), which is power-bounded by Allan (1997). Burlando (1997) proved that given $j \in \mathbb{N}$, if T on a complex X satisfies $n^{-j}\|T^n\| \rightarrow 0$ and 1 is a pole of the resolvent, then $n^{-j} \sum_{k=0}^{n-1} T^k$ converges in operator norm, and identified the limit; other conditions are also presented.

If T is uniformly ergodic on X real or complex, then $n^{-1}\|T^n\| \rightarrow 0$ and $(I - T)X$ is necessarily closed. Lin (1974a) proved the following (which is easy, using (2), when T is already mean ergodic, e.g., X is reflexive and T power-bounded).

Theorem 23 *Let T on a (real or complex) Banach space X satisfy $n^{-1}\|T^n\| \rightarrow 0$. Then T is uniformly ergodic if and only if $(I - T)X$ is closed.*

As a corollary, it is shown that also in the real case, a power-bounded quasi-compact operator is uniformly ergodic. Uniform ergodicity then implies that $I - T$ is invertible on $(I - T)X$, hence $(I - T)^k X = (I - T)X$ is closed for any $k \geq 1$. Mbekhta and Zemánek (1993) showed that when X is complex, one can replace condition (iii) in Theorem 22 by (iii)': $(I - T)^k X$ is closed for some $k \geq 1$. Grabiner and Zemánek (2002) proved that if T is Cesàro bounded and $(I - T)^k X$ is closed for some $k \geq 2$, then $X = \mathcal{F}(T) \oplus (I - T)X$.

Becker (2011) proved that if T is uniformly ergodic, then the one-sided ergodic Hilbert transform $Hx := \sum_{k=1}^{\infty} k^{-1} T^k x$ converges for every $x \in (I - T)X$. In Becker (2005), she proved that when $\|T^n\| = O(n^\alpha)$ for some $0 \leq \alpha < 1$, uniform ergodicity is equivalent to the set of convergence of H being closed.

Extending some results of Hille (1945), Yoshimoto (1998) proved the following.

Theorem 24 *Let T on a complex Banach space X satisfy $n^{-1}\|T^n\| \rightarrow 0$. Then T is uniformly ergodic with limit E if and only if $\|(\lambda - 1)R(\lambda, T) - E\| \rightarrow 0$ as $\lambda \rightarrow 1^+$.*

Improving upon some results of Koliha (1974), Luecke (1977) proved: (i) T^n converges in norm if and only if T is uniformly ergodic and $\|T^n(I - T)\| \rightarrow 0$. (ii) when X is complex, T^n converges in norm if and only if T is uniformly ergodic and $\sigma(T) \cap \mathbb{T} \subset \{1\}$. When T^n converges uniformly, the rate is exponential: $\|T^n - E\| \leq M\rho^n$ for some $M > 0$ and $\rho \in (0, 1)$.

Weiss (1989) proved the following discrete-time version of the Datko (1970) and Pazy (1972) theorem about exponential stability of C_0 -semigroups.

Theorem 25 *Let T be an operator on a Banach space X . Then $\|T^n\| \rightarrow 0$ if and only if for some (every) $p \in [1, \infty)$, $\sum_{n=0}^{\infty} \|T^n x\|^p < \infty$ for every $x \in X$.*

Recall that over $\mathbb{C}$ we have $\|T^n\| \rightarrow 0$ if and only if T has spectral radius $r(T) < 1$. Glück (2015) obtained the following improvement of Weiss' main result (in which p is constant).

Theorem 26 *Let T be a bounded operator on a complex Banach space X . Assume that for every $x \in X$ and $\varphi \in X^*$ there exists $p \in [1, \infty)$ such that $\sum_{n=0}^{\infty} |\varphi(T^n x)|^p < \infty$. Then T has spectral radius $r(T) < 1$.*

When a power-bounded T is uniformly ergodic, we have equality in both inclusions of (4). It was observed in Fonf et al. (1996) that a power-bounded T is uniformly ergodic if and only if $\left\{x \in X : \sup_n \left\| \sum_{k=0}^{n-1} T^k x \right\| < \infty\right\}$ is closed (and then equals $\overline{(I - T)X}$). Since power-bounded operators on reflexive spaces need not be uniformly ergodic, the equality

$$(I - T)X = \left\{x \in X : \sup_n \left\| \sum_{k=0}^{n-1} T^k x \right\| < \infty\right\} \quad (5)$$

does not imply uniform ergodicity. Fonf et al. (1996) studied the equality (5) and proved that if X is a separable real Banach space which does not contain an infinite-dimensional closed subspace isomorphic to a dual Banach space (e.g., c_0 , or more generally $X^* = \ell_1$), then (5) implies uniform ergodicity.

Lin (1978) proved that a uniformly ergodic positive operator on a Banach lattice, with $\mathcal{F}(T)$ finite-dimensional, is quasi-compact; this is not true in general (e.g., $T = -I$).

Lotz (1968) proved that when T is positive power-bounded on a complex Banach lattice, if $\lambda \in \sigma(T) \cap \mathbb{T}$, then all powers λ^k are in $\sigma(T)$. Hence if T is a positive power-bounded quasi-compact operator with $r(T) = 1$, there exists $d \geq 1$ such that every $\lambda \in \sigma(T) \cap \mathbb{T}$ (which is an eigenvalue by Theorem 21) is a d th root of unity. We thus obtain the following.

Proposition 27 *A positive power-bounded T on a complex Banach lattice is quasi-compact if and only if for some integer $d \geq 1$ and a finite-dimensional projection E we have $\|T^{nd} - E\| \rightarrow 0$.*

Proposition 27 was proved for contractions by Bartoszek (1988).

Lotz (1981) studied uniformly ergodic Markov-Feller operators on $C(K)$, and in Lotz (1985) he studied uniformly ergodic positive operators on L_∞ . Groh (1983) studied the peripheral spectrum of uniformly ergodic positive operators on C^* -algebras.

Strong Cesàro Convergence

Convergence of the averages $\frac{1}{n} \sum_{k=1}^n a_k \rightarrow a$ of a sequence $\{a_k\}_{k \geq 1}$ can sometimes be strengthened to $\frac{1}{n} \sum_{k=1}^n |a_k - a| \rightarrow 0$ (strong Cesàro convergence). It seems to have first appeared in Hardy and Littlewood (1913). Koopman and von Neumann (1932) essentially proved that a bounded sequence $\{a_k\}$ is strongly Cesàro convergent to a if and only if there exists a set $J \subset \mathbb{N}$ of density zero, i.e., satisfying $n^{-1} |J \cap \{1, 2, \dots, n\}| \rightarrow 0$, such that $\lim_{J \ni k \rightarrow \infty} a_k = a$.

A restatement of the theorem of Koopman and von Neumann (1932) is: Let T be a unitary operator on a complex Hilbert space $\mathcal{H}$. Then

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n |\langle x, y \rangle| &\rightarrow 0 \quad \forall y \in \mathcal{H} \\ \text{if and only if } \langle x, z \rangle &= 0 \\ \text{whenever } T^*z &= \lambda z, \quad |\lambda| = 1. \end{aligned} \quad (6)$$

It follows from the work of Foguel (1963) that if T is a contraction of a complex Hilbert space $\mathcal{H}$ and $0 \neq x \in \mathcal{H}$ satisfies $\lim_n |\langle T^n x, x \rangle| = \|x\|^2$, then $Tx = \lambda x$ for some $\lambda \in \mathbb{T}$. This result was explicitly stated and proved (differently) for contraction C_0 -semigroups in $\mathcal{H}$ by Goldstein (1993), and extended to Banach spaces by Goldstein and Nagy (1995).

An ergodic probability-preserving transformation whose Koopman operator T on the complex L_2 has no unimodular eigenvalues different from 1 is called *weakly mixing*; equivalently, by (6), the

transformation is weakly mixing if $\langle T^n f, g \rangle$ is strong Cesàro convergent to $\langle Ef, g \rangle$ for every $f, g \in L_2(\mu)$. If $T^n f \rightarrow Ef$ weakly for every $f \in L_2$, the transformation is called *mixing*. The category theorems of Halmos (1944) and Rokhlin (1948) show that there are weakly mixing transformations which are not mixing; operator theoretic analogues were proved by Eisner and Serény (2008). Chacon (1969) was the first to construct a weakly mixing transformation which is not mixing.

L. Jones and Lin (1976) proved the following:

Theorem 28 *Let T be power-bounded on a (real or complex) Banach space X . The following are equivalent for $x \in X$:*

- (i) $\frac{1}{n} \sum_{k=1}^n |\langle T^k x, \varphi \rangle| \rightarrow 0$ for every $\varphi \in X^*$.
- (ii) $\sup_{\varphi \in X^*, \|\varphi\| \leq 1} \frac{1}{n} \sum_{k=1}^n |\langle T^k x, \varphi \rangle| \rightarrow 0$.
- (iii) $\left\| \frac{1}{n} \sum_{j=1}^n T^{k_j} x \right\| \rightarrow 0$ for every increasing k_j with $\sup k_j/j < \infty$.

They proved that if $T^{n_i} x \rightarrow 0$ weakly for some $\{n_i\}$, then the conditions of Theorem 28 hold, and when X^* is separable (but not in general) such convergence is also necessary. The relationship of (i) and (iii) was discussed earlier in L. Jones (1971) and Nagel (1974). Additional conditions were studied by Hiai (1978). When T is power-bounded on a reflexive space X , (i) implies the existence of a set $J \subset \mathbb{N}$ of density zero such that weak limit of $\{T^n x\}_{J \ni n \rightarrow \infty}$ is zero: By (iii), we can assume X separable (spanned by the orbit of x), and then by reflexivity X^* is separable.

Zsidó (2007) proved that if we replace $\{T^k x\}$ by an arbitrary-bounded sequence $\{x_k\} \subset X$, (ii) and (iii) of Theorem 28 are still equivalent, but are not implied by (i). Condition (iii) for sequences in Hilbert space was studied by Berend and Bergelson (1986). Kunszenti-Kovács (2015) showed that when T is a contraction on a Hilbert space $\mathcal{H}$ and $x \in \mathcal{H}$ satisfies (i) of Theorem 28, then for every nonconstant polynomial q mapping $\mathbb{N}$ into itself there is a set J of density zero such that $\text{weak-} \lim_{J \ni j \rightarrow \infty} T^{q(j)} x = 0$.

When X is a complex Banach space, we would like to extend the characterization of condition (i) of Theorem 28, given in (6) for unitary operators on Hilbert spaces. Extending the work of Jacobs (1957), deLeeuw and Glicksberg (1961) studied weakly almost periodic semigroups of operators (the image of every vector under the semigroup is weakly conditionally compact), which are necessarily bounded. A very special case of their results is the following decomposition.

Theorem 29 *Let T be weakly almost periodic on a complex Banach space X . Then*

$$X = \overline{\text{span}}\{y : Ty = \lambda y, |\lambda| = 1\} \\ \oplus \{z : 0 \in \text{weak-closure}\{T^n z\}_{n \geq 0}\}. \quad (7)$$

Moreover, $0 \in \text{weak-closure}\{T^n x\}_{n \geq 0}$ if and only if $\langle x, \varphi \rangle = 0$ for every $\varphi \in X^*$ with $T^* \varphi = \lambda \varphi, |\lambda| = 1$.

Since, for T power-bounded, $\|T^n z\| \rightarrow 0$ if (and only if) $\|T^{n_k} z\| \rightarrow 0$ along some subsequence $\{n_k\}$, we obtain:

Corollary 30 *Let T be almost periodic on a complex Banach space X . Then*

$$X = \overline{\text{span}}\{y : Ty = \lambda y, |\lambda| = 1\} \\ \oplus \{z : \|T^n z\| \rightarrow 0\}.$$

Proposition 31 *Let T be weakly almost periodic on a complex Banach space X . Then*

$$\sup_{\|\varphi\| \leq 1} \frac{1}{n} \sum_{k=1}^n |\langle T^k x, \varphi \rangle| \rightarrow 0 \\ \text{if and only if } \langle x, \psi \rangle = 0 \\ \text{whenever } T^* \psi = \lambda \psi, |\lambda| = 1. \quad (8)$$

L. Jones and Lin (1980) proved (8) for T power-bounded on complex X , assuming only that $\{T^n x\}$ is weak-* sequentially compact in X^{**} (e.g., X does not contain an isomorphic copy of ℓ_1). Their proof uses the result for isometries in Hilbert spaces, proved in Krengel (1972), but does not use the DeLeeuw-Glicksberg theory. It is also

noted there that for general contractions the “if” part of (8) need not hold.

In connection with Theorem 28(iii), Eisner and Müller (2021) studied particular subsequences along which the averages of any weakly almost periodic operator, which has no unimodular eigenvalues, converge in norm.

DeLaubenfels and Vü (1996) proved the following (see also Theorem 45 below).

Theorem 32 *Let T be power-bounded on X over $\mathbb{C}$, with $\sigma(T) \cap \mathbb{T}$ countable. Then $\|T^n x\| \rightarrow 0$ if (and only if) $\langle x, \psi \rangle = 0$ whenever $T^* \psi = \lambda \psi, |\lambda| = 1$.*

Extending the result of Krengel (1972) for unitary operators, Müller and Tomilov (2010) proved:

Theorem 33 *Let T be power-bounded on a complex Hilbert space $\mathcal{H}$ with $\sigma(T) \cap \mathbb{T}$ infinite and no unimodular eigenvalues. Then there exists a dense set of vectors $x \in \mathcal{H}$ such that for some increasing $\{n_k\}$ we have $\langle T^{n_k} x, T^{n_j} x \rangle = 0$ when $j \neq k$.*

When T is power-bounded on X separable (over $\mathbb{C}$), the set of unimodular eigenvalues $\sigma_p(T) \cap \mathbb{T}$ is countable, by Jamison (1965), but $\sigma_p(T)$ may be uncountable (as the left shift on ℓ_2 shows).

Weak Stability and Mixing

Recall that an ergodic probability-preserving transformation is called *mixing* if its Koopman operator T on L_2 satisfies $T^n \rightarrow E$ in the weak operator topology. We therefore call a power-bounded T on a Banach space X weakly stable if for some E (necessarily a projection) $T^n \rightarrow E$ in the weak operator topology. Weak stability implies power-boundedness and mean ergodicity, so when X is over $\mathbb{C}$ the spectral radius of T does not exceed 1.

Theorem 34 (Foguel (1963)). *Let T be a contraction on a Hilbert space $\mathcal{H}$ and $x \in \mathcal{H}$. Then $T^n x \rightarrow 0$ weakly if and only if $\langle T^n x, x \rangle \rightarrow 0$. Hence $T^n x \rightarrow 0$ weakly if and only if $T^{*n} x \rightarrow 0$ weakly.*

The following result of Foguel (1963) shows that weak stability of contractions in $\mathcal{H}$ is a problem only for unitary operators.

Theorem 35 *Let T be a contraction on a Hilbert space $\mathcal{H}$ and let $\mathcal{K} := \{y \in \mathcal{H} : \|T^n y\| = \|T^{*n} y\| = \|y\| \ \forall n \geq 1\}$. Then $\mathcal{K}$ is invariant under T and T^* , the restriction $T|_{\mathcal{K}}$ is unitary, and $\lim \langle T^n z, z \rangle = 0$ for any $z \perp \mathcal{K}$.*

Horowitz (1969) proved that if in Theorem 35 T is normal, then $\|T^n z\| \rightarrow 0$ for $z \perp \mathcal{K}$.

In the following, Badea and Müller (2009) showed that in general the convergence to 0 of $\langle T^n x, x \rangle$ can be arbitrarily slow.

Theorem 36 *Let T on a complex Hilbert space $\mathcal{H}$ satisfy $T^n x \rightarrow 0$ weakly for every $x \in \mathcal{H}$ (so T is power-bounded), and assume $1 \in \sigma(T)$. Then for every positive sequence $r_n \rightarrow 0$ with $\sup_n r_n < 1$ there exists $x \in \mathcal{H}$ with $\|x\| = 1$ such that $\operatorname{Re} \langle T^n x, x \rangle > r_n$.*

The following theorem, proved independently by Jones and Kuftinec (1971) and by Akcoglu and Sucheston (1972), was first proved by Blum and Hanson (1960) for the L_2 Koopman operator of a mixing probability-preserving transformation.

Theorem 37 *Let T be a contraction in a Hilbert space $\mathcal{H}$ and $x \in \mathcal{H}$. Then*

$$\begin{aligned} T^n x &\rightarrow 0 \text{ weakly if and only if} \\ \left\| \frac{1}{n} \sum_{j=1}^n T^{k_j} x \right\| &\rightarrow 0 \\ \text{for every increasing } \{k_j\}. \end{aligned} \quad (9)$$

Krengel (1971) and Conze (1973) constructed increasing sequences $\{k_j\}$ such that for every ergodic mpt θ there exists $f \in L_1$ for which $\frac{1}{n} \sum_{j=1}^n f \circ \theta^{k_j}$ does not converge a.e.

Berend and Bergelson (1986) abstracted Theorem 37, by proving that for $\{x_n\} \subset \mathcal{H}$ bounded, $x_n \rightarrow 0$ weakly if and only if $\left\| \frac{1}{n} \sum_{j=1}^n x_{k_j} \right\| \rightarrow 0$ for every increasing $\{k_j\}$.

We say that a power-bounded T on a Banach space X has the *Blum-Hanson (BH) property* if $T^n x \rightarrow 0$ weakly implies $\left\| \frac{1}{n} \sum_{j=1}^n T^{k_j} x \right\| \rightarrow 0$ for

every increasing $\{k_j\}$; the converse implication always holds. Theorem 37 says that any contraction on a Hilbert space has the BH property. A weakly stable T on X , with $T^n \rightarrow E$ in the weak operator topology, has the BH property if $\left\| \frac{1}{n} \sum_{j=1}^n T^{k_j} x - Ex \right\| \rightarrow 0$ for every increasing $\{k_j\}$ and every $x \in X$. Theorem 37 raises the question whether any weakly stable operator on a Banach space has the BH property.

Akcoglu and Sucheston (1972) proved that weakly stable contractions on L_1 have the BH property. Millet (1976) proved that a weakly stable positive operator on L_1 has the BH property. On the other hand, Akcoglu et al. (1974) constructed a topological dynamical system (K, τ) such that its Koopman operator T on $C(K)$ is weakly stable, but for some $x \in C(K)$ (9) fails. Lefèvre and Matheron (2016) proved that when K is a compact metric space, all contractions have the BH property if and only if K has only finitely many accumulation points.

Akcoglu and Sucheston (1975), and then Bellows (1975), proved that any weakly stable positive contraction on L_p , $1 < p < \infty$, has the BH property. Müller and Tomilov (2007) proved that when T is a contraction of ℓ_p , $1 \leq p < \infty$, the equivalence (9) holds; they also showed that a power-bounded T on a Hilbert space may be weakly stable without possessing the BH property. Satō (1980) proved that if T is positive and weakly stable on L_p of a finite measure space, $1 < p < \infty$, and for some $p < p_1 < \infty$ T is also power-bounded in L_{p_1} , then T has the BH property on L_p .

Lefèvre et al. (2016) studied the BH property in terms of general sequences $\{x_n\} \subset X$ converging weakly to zero. They have additional examples of real Banach spaces in which every contraction has the BH property; see also Grivaux (2019). Azizov and Chilin (2017) proved the BH property for contractions on certain separable Banach lattices.

Eisner (2010) proved the following discrete analogue of a result of Chill and Tomilov (2007); when its condition (10) holds for any $x \in X$ and $\varphi \in X^*$, we obtain weak stability.

Theorem 38 *Let T on a complex Banach space have spectral radius $r(T) \leq 1$. If $x \in X$ and $\varphi \in X^*$ satisfy*

$$\lim_{r \rightarrow 1+} (r-1) \int_0^{2\pi} |\varphi(R(re^{it}, T)^2 x)| dt = 0, \quad (10)$$

then $\varphi(T^n x) \rightarrow 0$.

Leonov (1961), independently, proved Browder's Theorem 14 for weakly stable isometries in Hilbert space and showed that if T is the Koopman operator of an ergodic probability-preserving transformation on (Ω, Σ, μ) , then for any nontrivial $A \in \Sigma$ the function $f := 1_A - \mu(A)$ satisfies $\|\sum_{k=1}^n T^k f\|_2 \rightarrow \infty$ (so f is not a coboundary); see El Abdalaoui et al. (2010) for a related result.

Stability

A power-bounded operator T on a Banach space X is called *stable* if T^n converges in the strong operator topology, necessarily to a projection on $\mathcal{F}(T)$. A weakly stable T is stable if and only if it is almost periodic. Corollary 30 yields:

Proposition 39 *An operator T on a complex Banach space is stable if and only if it is almost periodic with no unimodular eigenvalues different from 1.*

Since Koopman operators of probability-preserving transformations are isometries in L_p , $1 \leq p < \infty$, they are never stable, unless $T = I$. However, in the noninvertible case the dual T^* of a Koopman operator may be stable, e.g., T induced by $\theta t = 2t \bmod 1$ on $[0, 1)$, by Horowitz (1968); this example shows that a contraction (i.e., T^*) on $\mathcal{H}$ may be stable while its dual is not. A simpler example is the shift $S(a_1, a_2, \dots) = (a_2, a_3, \dots)$ on $\ell_2(\mathbb{N})$, which is stable, while S^* is not, being an isometry.

In his study of stability of Markov operators, Derriennic (1976) proved the following.

Theorem 40 *Let T be a contraction of a Banach space X , and let B^* be the unit ball of the dual space X^* . Then for any $x \in X$ we have:*

- (i) $\lim_{n \rightarrow \infty} \|T^n x\| = \sup \{ |\langle x, \varphi \rangle| : \varphi \in \cap_{n \geq 1} T^{*n} B^* \}.$
- (ii) $\lim_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{k=1}^n T^k x \right\| = \sup \{ |\langle x, \varphi \rangle| : \varphi \in B^* \cap \mathcal{F}(T^*) \}.$

*Hence a mean ergodic contraction is stable if and only if $\cap_{n \geq 1} T^{*n} B^* \subset \mathcal{F}(T^*)$.*

Rokhlin (1961) called a probability-preserving transformation θ on (Ω, Σ, μ) exact if $\cap_{n \geq 1} \{\theta^{-n}(A) : A \in \Sigma\}$ is trivial modulo μ . Theorem 40 yields that θ is exact if and only if the dual of its Koopman operator is stable in L_2 , with limit $Ef = \int f d\mu$.

Lemma 41 *A mean ergodic operator T is (weakly) stable if and only if it is power-bounded and $T^n(I - T)x \rightarrow 0$ (weakly) for every $x \in X$.*

Lemma 41 is a direct consequence of (2). Here is an example for Lemma 41: Let S be power-bounded on a (real or complex) Banach space, and for $0 < \alpha < 1$ put $T = (1 - \alpha)I + \alpha S$. Since $I - T = \alpha(I - S)$, by Foguel and Weiss (1973) we have $\|T^n(I - T)\| \leq K/\sqrt{n} \rightarrow 0$; thus, T is stable when S is mean ergodic. In the complex case, any S with $r(S) = 1$ is the uniform limit of the uniformly stable operators $S_n := (1 - \frac{1}{n})S$.

Katznelson and Tzafriri (1986) proved the following deep result.

Theorem 42 *Let T be a power-bounded operator on a complex Banach space. Then $\|T^n(I - T)\| \rightarrow 0$ if and only if $\sigma(T) \cap \mathbb{T} \subset \{1\}$.*

The case $\sigma(T) = \{1\}$ was proved by Esterle (1983). Allan et al. (1987) showed that $\sigma(T) \cap \mathbb{T} \subset \{1\}$ implies the norm convergence of $\sum_{n \geq 0} \zeta^n (T^n - T^{n+1})$ for any $1 \neq \zeta \in \mathbb{T}$. Allan and Ransford (1989) proved that Theorem 42 is equivalent to an old result of Gelfand (1941), which says that if T is invertible with $\sup_{n \in \mathbb{Z}} \|T^n\| < \infty$ and $\sigma(T) = \{1\}$, then $T = I$. A different proof of Theorem 42 was given by Vù (1992a). For additional historical information, we refer to Zemánek (1994). While Gelfand's result is valid also if $M_n(T)$ and $M_n(T^{-1})$ are bounded, by Drissi and Zemánek (2000), Tomilov and Zemánek (2004) showed that Theorem 42 fails if power-boundedness is replaced by Cesàro

boundedness (or even mean ergodicity); Léka (2010) has an example with $\sigma(T) = \{1\}$. For additional information and new proofs see the survey of Batty and Seifert (2022).

Allan et al. (1987) proved the following.

Theorem 43 *Let $f(z) = \sum_{k=0}^{\infty} a_k z^k$ with $\sum_{k \geq 0} k |a_k| < \infty$, and let T be power-bounded on a complex Banach space X . Then $\|T^n f(T)\| \rightarrow 0$ if and only if $f(\lambda) = 0$ for every $\lambda \in \sigma(T) \cap \mathbb{T}$.*

The conclusion of Theorem 43 was proved by Esterle et al. (1990) for every f in the disk algebra when T is a contraction of a Hilbert space. When X is a Hilbert space, Léka (2009) extended Theorem 43 to f with $\sum_k |a_k| < \infty$. Huang (1995) proved that if $\|T^n f(T)\| \rightarrow 0$ for some $f(z)$ with $\sum_k |a_k| < \infty$, then $\sigma(T) \cap \mathbb{T}$ has measure zero. A continuous time version of the Katznelson-Tzafriri theorem, for bounded C_0 -semigroups, was proved by Esterle et al. (1992) and by Vũ (1992b).

Following Kalton et al. (2004), see Theorem 67 below, Malinen et al. (2007) proved that for any T , either $\liminf_n (n+1) \|T^n(I-T)\| = 0$, or $\liminf_n (n+1) \|T^n(I-T)\| \geq e^{-1}$. Dungey (2008) gave several necessary and sufficient conditions for $\|T^n(I-T)\| = O(n^{-1/2})$ in Theorem 42. The rate $1/n$ is related to the Ritt resolvent condition; see the section *Resolvent conditions and growth of powers* below. Seifert (2016) obtained a rate in Theorem 42 using the growth of $\|R(e^{it}, T)\|$ as $t \rightarrow 0$. Cohen and Lin (2016) studied the rate $1/n^\alpha$, $0 < \alpha \leq 1$, when X is a Hilbert space. Badea and Seifert (2017) obtained a condition on the numerical range which implies $\sigma(T) \cap \mathbb{T} \subset \{1\}$. Ng and Seifert (2020) got the full correspondence, when X is a Hilbert space, between the rate of growth of $\|R(e^{it}, T)\|$ as $t \rightarrow 0$ and the rate of $\|T^n(I-T)\| \rightarrow 0$.

Baskakov (2015) obtained a representation of T^n when $\sigma(T) \cap \mathbb{T}$ is finite, an immediate corollary of which is Theorem 42.

A contraction T on a Hilbert space $\mathcal{H}$ is called *completely nonunitary* if its unitary subspace $\mathcal{K}$, defined in Theorem 35, is trivial. In that case, $T^n x \rightarrow 0$ weakly for every $x \in \mathcal{H}$, by Theorems

35 and 34. In Sect. II.6 of their book, Sz-Nagy and Foiaş (1967) proved the following.

Theorem 44 *Let T be a completely nonunitary contraction on a complex Hilbert space $\mathcal{H}$. If $\sigma(T) \cap \mathbb{T}$ has zero Haar measure (on $\mathbb{T}$), then $\|T^n x\| \rightarrow 0$ and $\|T^{*n} x\| \rightarrow 0$ for any $x \in \mathcal{H}$.*

Following Kérchy and van Neerven (1997), Mustafayev (2014) studied conditions on the local spectrum which imply local stability. His work yields a partial extension of Theorem 44 to power-bounded operators, by restricting the null-sets containing $\sigma(T) \cap \mathbb{T}$.

For the general case, Arendt and Batty (1988) proved the following; alternate proofs are indicated in Batty (1994). Lyubich and Vũ (1988) proved the continuous case, and their proof was adapted to the discrete case in Eisner (2010). See also Batty and Vũ (1990).

Theorem 45 *Let T be a power-bounded operator on a complex Banach space X . If $\sigma(T) \cap \mathbb{T}$ is countable and T^* has no unimodular eigenvalues, then $\|T^n x\| \rightarrow 0$ for $x \in X$.*

Eisner (2010) proved a discrete analogue of a resolvent condition of Tomilov (2001) which yields stability, and obtained as a corollary the following.

Theorem 46 *Let T be power-bounded on a complex Hilbert space $\mathcal{H}$ and $x \in \mathcal{H}$. Then $\|T^n x\| \rightarrow 0$ if and only if*

$$\lim_{r \rightarrow 1+} \int_0^{2\pi} \|R(re^{it}, T)x\|^2 dt = 0$$

Lin (1998) used Theorem 42 to prove the following.

Theorem 47 *Let T be a positive power-bounded operator on a Banach lattice. Then $\|T^n(I-T)\| \rightarrow 0$ if and only if $\inf_n \|T^n(I-T)\| < \sqrt{3}$.*

For contractions on L_p , $2 \neq p \in (1, \infty)$, the constant $\sqrt{3}$ can be replaced by a constant $\sqrt{3} < c_p < 2$, computed in Berend (1992). For L_1 , the constant is 2, by Foguel (1976).

Following Lasota et al. (1984), Sine (1991) defined a power-bounded operator T on a Banach space X to be *constrictive* if there exists a compact set $K \subset X$ such that $\text{dist}(T^n x, K) \rightarrow 0$ for every $\|x\| \leq 1$. When $K = \{0\}$, we have stability. By the definition, if T is constrictive, then for each $x \in X$ the orbit $\{T^n x\}$ is precompact, so T is almost periodic, hence mean ergodic. Sine (1991) proved that a contraction T is constrictive if and only if T is strongly almost periodic and, in the decomposition (7), $\overline{\text{span}}\{y : Ty = \lambda y, |\lambda| = 1\}$ is finite-dimensional (and on its complement T is stable). Extending the result of Lasota et al. (1984) for L_1 , Bartoszek (1988) proved the following.

Theorem 48 *Let T be a constrictive positive contraction of a Banach lattice. Then there exist r positive unit vectors $y_1, \dots, y_r$ and r positive functionals $\varphi_1, \dots, \varphi_r$ such that T permutes the y_j , and $\left\|T^n \left(x - \sum_{j=1}^r \varphi_j(x) y_j\right)\right\| \rightarrow 0$ for every x . Hence, for some $1 \leq d \leq r$, T^d is stable.*

Averaging Along Subsequences

The Blum-Hanson Theorem 37 and Theorem 28 (iii) raise the question for which increasing subsequences of positive integers $\{k_j\}$ the averages $\frac{1}{n} \sum_{j=1}^n T^{k_j}$ converge strongly for contractions, or power-bounded operators, on a Hilbert space $\mathcal{H}$. Blum and Eisenberg (1974) proved a “uniform distribution” criterion on a sequence $\{k_j\}$ for the strong convergence of $\frac{1}{n} \sum_{j=1}^n T^{k_j}$ to the ergodic limit $E(T)$ for every unitary operator; it yields also convergence for any contraction on $\mathcal{H}$, by the unitary dilation theorem. For contractions on $\mathcal{H}$, the following general criterion, implied in Furstenberg (1981) for unitary operators, was presented for the complex case in Rosenblatt (1994), in Rosenblatt and Wierdl (1995), and in Berend et al. (2002). For the real case, one applies the complex case in the complexification of $\mathcal{H}$, as in Michal and Wyman (1941).

Proposition 49 *Let $\{k_j\}$ be an increasing sequence of positive integers. The averages $\frac{1}{n} \sum_{j=1}^n T^{k_j} x$ converge in norm for every*

contraction on a (real or complex) Hilbert space $\mathcal{H}$ and $x \in \mathcal{H}$ if and only if for every $\lambda \in \mathbb{T}$ the averages $\frac{1}{n} \sum_{j=1}^n \lambda^{k_j}$ converge.

Motivated by a result of Brunel and Keane (1969), Bourgain (1988b) and Bourgain et al. (1989) proved the “return times theorem,” which provides increasing sequences $\{k_j\}$, with $\lim k_j/j < \infty$, such that averaging any Koopman operator on any L_p along $\{k_j\}$ converges a.e., hence also in L_p norm ($1 \leq p < \infty$). Applying it to all rotations of $\mathbb{T}$, we obtain, by Proposition 49, that for these sequences $\frac{1}{n} \sum_{j=1}^n T^{k_j} x$ converges in norm for any contraction T on a Hilbert space $\mathcal{H}$ and $x \in \mathcal{H}$.

Extending a result of Furstenberg (1981) for unitary operators (for which Bergelson (1987) gave a nonspectral proof), Kunszenti-Kovács et al. (2011) proved the following.

Theorem 50 *Let q be a nonconstant real polynomial with integer coefficients which maps $\mathbb{N}$ to itself. Then the limit $c_q(\lambda) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \lambda^{q(k)}$ exists for $\lambda \in \mathbb{T}$ and is zero for λ not a root of unity.*

For every contraction T on a Hilbert space $\mathcal{H}$ and $x \in \mathcal{H}$, the averages $\frac{1}{n} \sum_{k=1}^n T^{q(k)} x$ converge in norm.

If $\mathcal{H}$ is over $\mathbb{C}$, then the limit is $\sum_{\lambda: \exists j \lambda^j = 1} c_q(\lambda) E_\lambda x$, where $E_\lambda x := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} (\bar{\lambda} T)^k x$ is the orthogonal projection on the eigenspace corresponding to $\lambda \in \mathbb{T}$.

Given a contraction T on $\mathcal{H}$, the limit is a projection for every q with $q(0) = 0$ if and only if roots of unity different from ± 1 are not eigenvalues of T .

Proposition 49 does not apply to power-bounded operators. Using a different method, ter Elst and Müller (2017) proved:

Theorem 51 *Let q be a nonconstant real polynomial with rational coefficients which maps $\mathbb{N}$ to itself. If T is a power-bounded operator on Hilbert space $\mathcal{H}$ over $\mathbb{C}$ such that roots of unity are not eigenvalues, then $\frac{1}{n} \sum_{k=1}^n T^{q(k)} x \rightarrow 0$ for every $x \in \mathcal{H}$.*

Eisner and Müller (2021) extended Theorem 51 by proving that for a general real polynomial q with $[q(k)] \geq 0$ for $k \in \mathbb{N}$, we have convergence of $\frac{1}{n} \sum_{k=1}^n T^{[q(k)]} x$ for every $x \in \mathcal{H}$.

Bourgain (1989) proved that for T the Koopman operator on L_p of an invertible probability-preserving transformation, we have a.e. convergence of $\frac{1}{n} \sum_{k=1}^n T^{q(k)} f$ for $f \in L_p$, $1 < p < \infty$, and q as in Theorem 51. This yields L_p -norm convergence, also in L_1 . For general real polynomials, he proved a.e. convergence of $\frac{1}{n} \sum_{k=1}^n T^{[q(k)]} f$ for $f \in L_\infty$.

Let $\{p_j\}$ be the sequence of prime numbers in increasing order. Extending earlier results of Bourgain (1988a) and of Wierdl (1988), Nair (1993) proved that given q a nonconstant polynomial mapping $\mathbb{N}$ to itself, for the Koopman operator as above on L_p , $1 < p < \infty$, we have a.e. convergence of $\frac{1}{n} \sum_{j=1}^n T^{q(p_j)} f$ for $f \in L_p$, hence in L_p norm. Applying Nair's result to all rotations of $\mathbb{T}$ and using Proposition 49, we obtain the following result, proved directly in Eisner and Lin (2018).

Theorem 52 *Let $\{p_j\}$ be the sequence of primes in increasing order. Given a nonconstant polynomial q mapping $\mathbb{N}$ to itself, for every contraction on a (real or complex) Hilbert space $\mathcal{H}$ and $x \in \mathcal{H}$, the averages $\frac{1}{n} \sum_{j=1}^n T^{q(p_j)} x$ converge in norm.*

Boshernitzan and Wierdl (1996) studied conditions on certain real sequences $a(t)$, which tend to $+\infty$ as $t \rightarrow \infty$, such that for every Koopman operator T of an invertible probability-preserving transformation, the averages $\frac{1}{n} \sum_{k=1}^n T^{[a(k)]} f$ converge in L_2 -norm.

General Averaging Methods

Since the averages of a convergent sequence converge to the same limit, it is natural to apply more general averaging (summability) methods which preserve convergence and limits, when convergence of the arithmetic means fails.

A matrix $\mathbf{A} = (a_{n,j})_{n,j \geq 0}$ maps ℓ_∞ into itself, by $(\mathbf{A}\mathbf{b})_n = \sum_{j=0}^\infty a_{n,j} b_j$ for $\mathbf{b} := (b_0, b_1, \dots) \in \ell_\infty$, if and only if $\sup_n \sum_{j=0}^\infty |a_{n,j}| < \infty$; if, in addition, the matrix $\mathbf{A}$ satisfies $\lim_{n \rightarrow \infty} a_{n,j} = 0$ for every j and $\lim_{n \rightarrow \infty} \sum_{j=0}^\infty a_{n,j} = 1$, then it maps the space c of convergent sequences into itself, with

preservation of the limit (Silverman-Toeplitz theorem). Such a matrix is called a *regular summability matrix*, or in short a *regular matrix*. The Cesàro averaging corresponds to the matrix with $a_{n,j} = 1/(n+1)$ for $j \leq n$ and 0 for $j > n$. The first to consider averaging of powers of a power-bounded operator by regular matrices was L. Cohen (1940), who proved:

Theorem 53 *Let $\mathbf{A} = (a_{n,j})_{n,j \geq 0}$ be a regular matrix such that*

$$\lim_{k \rightarrow \infty} \sum_{j=k}^\infty |a_{n,j+1} - a_{n,j}| = 0 \text{ uniformly in } n.$$

If T is a power-bounded operator on a Banach space X such that for every $x \in X$ the sequence $B_n x := \sum_{j=0}^\infty a_{n,j} T^j x$ is weakly conditionally compact, then for every x the sequence $B_n x$ converges to a fixed point.

In particular, when all the elements of $\mathbf{A}$ above are nonnegative with $\sum_{j=0}^\infty a_{n,j} = 1$ for every n , then the above convergence holds for every weakly almost periodic T .

Eberlein (1984) looked at the B_n in Theorem 53 as invariant integrals, in the sense of Eberlein (1949).

Note that when the regular matrix $\mathbf{A}$ is triangular (i.e., $a_{n,j} = 0$ for $j > n$), the series $B_n x := \sum_{j=0}^\infty a_{n,j} T^j x$ is a finite sum, so is defined for any operator T , and one can consider the convergence of $B_n x$ even for T not power-bounded.

Let $\{w_k\}_{k \geq 0}$ be a sequence of nonnegative numbers with $w_0 > 0$ and $\sum_{k=0}^\infty w_k = \infty$, and put $W_n = \sum_{k=0}^n w_k$. The matrix $\mathbf{A}$ defined by $a_{n,j} = w_j/W_n$ for $j \leq n$ and zero for $j > n$ is regular; it yields the *weighted averages* with weights $\{w_k\}$. Lin et al. (1999) proved:

Theorem 54 *Let $\{w_k\}$ be nonnegative with $w_0 > 0$ and $\sum_{k=0}^\infty w_k = \infty$. The weighted averages $\frac{1}{W_n} \sum_{k=0}^n w_k T^k x \rightarrow Ex$ for every mean ergodic T on a Banach space X and $x \in X$ if and only if*

$$\lim_{n \rightarrow \infty} \frac{1}{W_n} \left(w_n + \sum_{k=0}^{n-1} |w_{k+1} - w_k| \right) \rightarrow 0.$$

The latter condition is satisfied when $\{w_k\}$ is nondecreasing and w_n/W_n tends to 0.

For $\alpha > 0$, the Cesàro- α averages of an operator T , denoted $C_n^\alpha(T)$, are defined as follows: For $\beta > -1$, let $A_0^\beta = 1$ and $A_n^\beta := (\beta + 1) \cdots (\beta + n)/n!$ for $n > 0$, and let $C_n^\alpha(T) = \sum_{j=0}^n \left(A_{n-j}^{\alpha-1} / A_n^\alpha \right) T^j$. The corresponding matrix $a_{nj} := A_{n-j}^{\alpha-1} / A_n^\alpha$ for $0 \leq j \leq n$ and 0 for $j > n$ is regular and satisfies the conditions of Theorem 53. An operator T on X is called (weakly) (C, α) ergodic if $C_n^\alpha(T)x$ converges (weakly) as $n \rightarrow \infty$, for every $x \in X$. When $\sup_n \|C_n^\alpha(T)\| < \infty$, we say that T is (C, α) bounded. Note that $(C, 1)$ (weak) ergodicity is (weak) mean ergodicity.

Li et al. (2008) proved that for every $0 < \alpha < 1$ there exists a positive T on some L_1 such that T is (C, β) bounded for $\beta > \alpha$, but is not (C, α) bounded.

Let T satisfy $\limsup_{n \rightarrow \infty} \|T^n\|^{1/n} \leq 1$ (spectral radius $r(T) \leq 1$ when X is over $\mathbb{C}$). Then for $0 < r < 1$, the series $\sum_{k=0}^\infty r^k T^k$ converges in operator norm, and we call $A_r = A_r(T) := (1-r) \sum_{k=0}^\infty r^k T^k$ the Abel averages of T . For any sequence $0 < r_n \uparrow 1$, the matrix $a_{nj} = (1-r_n)r_n^j$ is regular and satisfies the assumptions of Theorem 53. An operator T as above is called Abel bounded if $\sup_{0 < r < 1} \|A_r(T)\| < \infty$, and Abel ergodic if $Ex := \lim_{r \rightarrow 1^-} A_r(T)x$ exists for every x ; in the latter case $Ex \in \mathcal{F}(T)$, and we have $X = \mathcal{F}(T) \oplus \overline{(I - T)X}$. If for T as above and $\lambda > 1$, we denote $R(\lambda, T) := \sum_{k=0}^\infty \lambda^{-(k+1)} T^k$ (the resolvent at λ in the complex case), then we have $\lim_{r \rightarrow 1^-} A_r(T)x = \lim_{\lambda \rightarrow 1^+} (\lambda - 1)R(\lambda, T)x$.

Hille (1945) studied (C, α) and Abel averaging of general sequences in a Banach space. The following is his application of his general results.

Theorem 55 *Let $\alpha > 0$. If T is a (C, α) ergodic operator on a Banach space X , then $\|T^n x\|/n^\alpha$ converges to zero for any $x \in X$, and T is Abel ergodic, with $\lim_{r \rightarrow 1^-} A_r(T)x = \lim_{n \rightarrow \infty} C_n^\alpha(T)x$.*

If T is power-bounded and Abel ergodic, then it is (C, α) ergodic.

If T is (C, α) bounded and Abel ergodic, it is (C, β) ergodic for any $\beta > \alpha$.

Analogous results for C_0 -semigroups were proved by Kendall and Reuter (1956), using the abstract ergodic theorems of Eberlein (1949).

Hille (1945) applied his results to the operator norm topology. The converse of the first statement in Theorem 55 was proved for the operator norm topology by Badiozzaman and Thorpe (1992) and reproved by Ed-dari (2003); put together, the following was obtained (proved by Yoshimoto (1998) for $0 < \alpha \leq 1$).

Theorem 56 *Let $\alpha > 0$. The operator T is uniformly (C, α) ergodic if and only if it is uniformly Abel ergodic and $n^{-\alpha} \|T^n\| \rightarrow 0$.*

An analogous result for the strong operator topology was proved in Ed-dari (2004).

Lin et al. (2015) proved that T is uniformly Abel ergodic if and only if for some (all) $0 < r < 1$, $(A_r(T))^n$ converges uniformly as $n \rightarrow \infty$; Kozitsky et al. (2013) gave a spectral condition for this property.

Li et al. (2008) showed that in any infinite-dimensional Banach space, Abel boundedness does not imply Cesàro boundedness. However, Émilion (1985) proved that any Abel bounded positive operator on a Banach lattice is Cesàro bounded, and when the lattice is reflexive it is mean ergodic.

Improving the extension of Gelfand's theorem by Mbekhta and Zemánek (1993), Grobler and Huijsmans (1995) proved that if $\sigma(T) = \{1\}$ and both T and T^{-1} are Abel bounded, then $T = I$.

A regular matrix $\mathbf{A}$ is called uniformly regular if $\lim_{n \rightarrow \infty} \sup_j |a_{nj}| = 0$. Fong and Sucheston (1974), extending an earlier result of Hanson and Pledger (1969), proved the following extension of the Blum-Hanson Theorem 37.

Theorem 57 *Let T be a contraction on a (real or complex) Hilbert space $\mathcal{H}$ and let $x \in \mathcal{H}$. Then $T^n x$ converges weakly if and only if for every uniformly regular matrix $\mathbf{A}$ the averages $B_n x = \sum_{j=0}^\infty a_{nj} T^j x$ converge in norm.*

They also proved that when T is a contraction of L_1 , the convergence of T^n in the weak operator topology is equivalent to the convergence in the strong operator topology of $\sum_{j=0}^\infty a_{nj} T^j$ for every uniformly regular matrix $\mathbf{A}$. Akcoglu and Sucheston (1975) proved the same equivalence

for positive contractions of L_p , $1 < p < \infty$. Lin and Weber (2007) used Theorem 57 to obtain:

Theorem 58 *Let $w_j \geq 0$ with $w_0 > 0$ and $W_n := \sum_{j=0}^n w_j \rightarrow \infty$. Then for every contraction T on a Hilbert space with $T^n \rightarrow 0$ weakly, the weighted averages $\frac{1}{W_n} \sum_{j=0}^n w_j T^j$ converge strongly if and only if $\frac{1}{W_n^2} \sum_{j=0}^n w_j^2 \rightarrow 0$.*

Dunford and Schwartz (1956) extended the pointwise ergodic theorem, proving that if T is a contraction of $L_1(\mu)$ of a σ -finite measure, such that $\|Tf\|_\infty \leq \|f\|_\infty$ for $f \in L_1 \cap L_\infty$, then for every $f \in L_1$ the averages $\frac{1}{n} \sum_{k=0}^{n-1} T^k f$ converge a.e. Akcoglu (1975) proved that if T is a positive contraction of $L_p(\mu)$, $1 < p < \infty$, then $\frac{1}{n} \sum_{k=0}^{n-1} T^k f$ converges a.e. for every $f \in L_p$. By classical summability theory for numerical sequences (extended in Hille (1945) to vector sequences), in both the above theorems $\lim_{r \rightarrow 1^-} A_r(T)f$ exists a.e. This raises the question if, when μ is finite, $C_n^\alpha(T)f$, $0 < \alpha < 1$ converges a.e. for every $f \in L_p$, in either theorem. Irmisch (1980) proved that in Akcoglu's theorem, the convergence holds when $\alpha > 1/p$, but may fail when $\alpha = 1/p$. D  niel (1989) proved Irmisch's result for the Koopman operator of a probability-preserving transformation and gave an example that even for such a Koopman operator the a.e. convergence may fail when $\alpha = 1/p$; the norm convergence holds for every $0 < \alpha < 1$ by Theorem 55.

Modulated Ergodic Theorems

Let T be a weakly almost periodic operator on a complex Banach space. Then for any $\lambda \in \mathbb{T}$ the operator λT is also weakly almost periodic, so by Theorem 7 $\frac{1}{n} \sum_{k=0}^{n-1} \lambda^k T^k$ converges in the strong operator topology. It follows that if $\mathbf{a} := \{a_k\}_{k \geq 0} \in \ell_\infty$ can be approximated uniformly in k by sequences $\mathbf{p} = \{p(k)\}$ of the form $p(k) := \sum_{j=1}^m c_j \lambda_j^k$ with $c_j \in \mathbb{C}$ and $\lambda_j \in \mathbb{T}$ (called *trigonometric polynomial sequences*), then the averages $\frac{1}{n} \sum_{k=0}^n a_k T^k$ converge in the strong operator topology.

Let T be the Koopman operator of a probability-preserving θ on (Ω, Σ, μ) , and fix

$\lambda \in \mathbb{T}$. By looking at the product of θ with the rotation by λ , acting on $(\Omega \times \mathbb{T})$, it can be deduced from Birkhoff's theorem that $\frac{1}{n} \sum_{k=0}^{n-1} \lambda^k T^k f$ converges a.e. for any $f \in L_1(\mu)$. Wiener and Wintner (1941) proved that when θ is ergodic, for $f \in L_1(\mu)$ there is a null set $A \in \Sigma$ such that for $\omega \notin A$, $\frac{1}{n} \sum_{k=0}^{n-1} \lambda^k T^k f(\omega)$ converges for every $\lambda \in \mathbb{T}$. The Wiener-Wintner theorem follows by applying the extension by Rudolph (1994) of the return times theorem to all rotations of $\mathbb{T}$; see the discussion following Proposition 65. See also Assani (2003). The Wiener-Wintner theorem yields that if $\mathbf{a} \in \ell_\infty$ can be uniformly approximated by trigonometric polynomials, then for every Koopman operator T and every $f \in L_1(\mu)$ the averages $\frac{1}{n} \sum_{k=0}^n a_k T^k f$ converge a.e.

Ryll-Nardzewski (1975) noted that if $\mathbf{k} = \{k_j\}_{j \geq 0}$ is an increasing sequence in $\mathbb{N}$ with $\lim_n k_n/n = d > 0$, then averaging along $\mathbf{k}$ is equivalent to the modulated average with the zero-one weights $a_m = 1$ if and only if $m \in \{k_j\}$.

For the study of sequences $\mathbf{a} = \{a_k\}_{k \geq 0}$ such that the modulated averages $\frac{1}{n} \sum_{k=0}^n a_k T^k f$ converge *almost everywhere* for every Koopman operator of an ergodic mpt and every $f \in L_p$ (some fixed $1 \leq p < \infty$), we refer the reader to Bellow and Losert (1985), Bourgain et al. (1989), Rudolph (1994), and to the book of Assani (2003). For some recent results, see Fan (2019).

Let T be mean ergodic on a complex Banach space. A complex sequence $\mathbf{a} := \{a_k\}_{k \geq 0}$ will be called a *good modulating sequence* for T if the modulated averages $M_n(T, \mathbf{a}) := \frac{1}{n} \sum_{k=0}^{n-1} a_k T^k$ converge in the strong operator topology; the convergence is called a *modulated ergodic theorem* (the a_k need not be positive, so we avoid the often used terminology of "weighted" ergodic theorems). If $\mathbf{a}$ is a good modulating sequence for the L_2 -Koopman operators of all rotations of $\mathbb{T}$, then we must have that $c(\lambda) := \lim_n \frac{1}{n} \sum_{k=0}^{n-1} a_k \bar{\lambda}^k$ exists for every $\lambda \in \mathbb{T}$; such a sequence is called a *Hartman sequence*, and by Kahane (1961) the set $\{\lambda \in \mathbb{T} : c(\lambda) \neq 0\}$ is at most countable. Berend et al. (2002) showed that a second necessary condition for modulating all rotations is that $\sup_n \sup_{\lambda \in \mathbb{T}} |\frac{1}{n} \sum_{k=0}^{n-1} a_k \lambda^k| < \infty$, and these two conditions together imply that $\mathbf{a}$ is a good

modulating sequence for every contraction T of a complex Hilbert space. The limit was identified in Lin et al. (1999):

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} a_k T^k x = \sum_{|\lambda|=1} c(\lambda) E(\bar{\lambda}, T)x, \quad x \in \mathcal{H},$$

where $E(\lambda, T)$ is the orthogonal projection on the eigenspace of λ (note that also $\{\lambda \in \mathbb{T} : E(\lambda, T)x \neq 0\}$ is countable).

For $1 \leq p < \infty$, denote $W_p := \{\mathbf{a} : \|\mathbf{a}\|_{W_p} := (\limsup_n \frac{1}{n} \sum_{k=0}^{n-1} |a_k|^p)^{1/p} < \infty\}$. Then $\|\mathbf{a}\|_{W_p}$ is a seminorm, and W_p is complete. We put $W_\infty := \ell_\infty$. Çömez et al. (1998) proved:

Theorem 59 Let $\mathbf{a} \in W_p$, $1 < p < \infty$ be Hartman. Then $\mathbf{a}$ is a good modulating sequence for every weakly almost periodic operator on a complex Banach space.

By Berend et al. (2002), Theorem 59 may fail when $\mathbf{a} \in W_1$. However, the necessity of $\mathbf{a} \in W_1$ follows from the next result, due to Lin et al. (1999).

Theorem 60 A sequence $\mathbf{a}$ is a good modulating sequence for every almost periodic operator on a complex Banach space if and only if it is Hartman and $\mathbf{a} \in W_1$.

Eisner and Lin (2018) proved the following.

Theorem 61 Let $\mathbf{a} \in W_1$. Then for every power-bounded T on a Banach space X and $x \in X$ with $T^n x \rightarrow 0$ (weakly), we have $\frac{1}{n} \sum_{k=0}^{n-1} a_k T^k x \rightarrow 0$ (weakly).

By Berend et al. (2002) there exists a Hartman sequence $\mathbf{a} \notin W_1$ which satisfies $\sup_n \sup_{\lambda \in \mathbb{T}} |\frac{1}{n} \sum_{k=0}^{n-1} a_k \lambda^k| < \infty$, so this $\mathbf{a}$ is a good modulating sequence for all contractions in a Hilbert space, but not for all almost periodic operators. Eisner and Lin (2018) proved:

Proposition 62 Let $\mathbf{a} \in W_1$. For every contraction T on a Hilbert space $\mathcal{H}$ and $x \in \mathcal{H}$ with $T^n x \rightarrow 0$ weakly, we have $\|\frac{1}{n} \sum_{k=1}^n a_k T^k x\| \rightarrow 0$ if and only if $a_n = o(n)$.

For mean ergodic operators, Lin et al. (1999) showed:

Theorem 63 A sequence $\mathbf{a} = \{a_k\}$ is a good modulating sequence for every mean ergodic power-bounded operator if and only if it is Hartman, $\mathbf{a} \in W_1$, and $\frac{1}{n} \sum_{k=0}^{n-1} |a_{k+1} - a_k| \rightarrow 0$.

A bounded sequence $\mathbf{a}$ is called (weakly) almost periodic if its orbit in ℓ_∞ under the shift $S(\{b_k\}) := \{b_{k+1}\}$ is conditionally (weakly) compact. Then the restriction of S to $\overline{\text{span}}\{S^k \mathbf{a}\}$ is (weakly) almost periodic. Eisner (2013) used the Jacobs-deLeeuw-Glicksberg decomposition (Theorem 29) to obtain the following.

Theorem 64 Let S be a weakly almost periodic operator on a complex Banach space Y , and fix $y \in Y$ and $\phi \in Y^*$. Then the sequence $\mathbf{a} := \{\phi(S^k y)\}_{k \geq 0}$ is Hartman (and obviously bounded). Hence a weakly almost periodic sequence is Hartman.

Lin et al. (1999) used the Wiener-Wintner theorem to generate Hartman sequences.

Proposition 65 Let τ be an ergodic probability-preserving transformation on $(\tilde{\Omega}, \tilde{\Sigma}, \nu)$, and fix $g \in L_p(\nu)$, $1 \leq p \leq \infty$. Then for a.e. $\tilde{\omega} \in \tilde{\Omega}$ the sequence $\mathbf{a} := \{g(\tau^k \tilde{\omega})\}_{k \geq 0}$ is Hartman and in W_p .

By Berend et al. (2002), the Hartman sequences $\mathbf{a} \in W_p$ defined in Proposition 65 are good modulating sequences for every contraction on a Hilbert space.

Rudolph (1994) extended Bourgain's return times theorem, showing that for $p \geq 1$, the Hartman sequences $\mathbf{a} \in W_p$ defined in Proposition 65 yield a.e. convergence of $\frac{1}{n} \sum_{k=0}^{n-1} a_k f(\theta^k \omega)$ for every probability-preserving θ on (Ω, Σ, μ) and $f \in L_q(\mu)$, $q = p/(p-1)$. Demeter et al. (2008) proved that when $p > 1$, the above sequences $\mathbf{a} \in W_p$ yield a.e. convergence of $\frac{1}{n} \sum_{k=0}^{n-1} a_k f(\theta^k \omega)$ for every probability preserving θ on (Ω, Σ, μ) and $f \in L_2(\mu)$; Demeter (2012) refined it to $f \in L_q$, $q^{-1} < \frac{3}{2} - p^{-1}$.

A sequence $\mathbf{a}$ which for some $1 \leq p < \infty$ is in the W_p closure of the trigonometric polynomial sequences is called a p -Besicovitch sequence. The

1-Besicovitch sequences are called Besicovitch sequences. Besicovitch sequences were first used by Tempelman (1974) and by Ryll-Nardzewski (1975) for obtaining pointwise modulated ergodic theorems for Koopman operators. Besicovitch sequences are Hartman, see Bellow and Losert (1985), and the space of 2-Besicovitch sequences is an inner product space. Parseval equality in this space, see Besicovitch (1954), yields uniqueness of the coefficient function $c(\lambda)$, i.e., $\mathbf{a}$ has $c(\lambda) \equiv 0$ if and only if $\frac{1}{n} \sum_{k=0}^n |a_k|^2 \rightarrow 0$. By Bellow and Losert (1985), all the sequences generated in Theorem 65 by τ invertible and $g \in L_\infty$ are Besicovitch if and only if τ has discrete spectrum (i.e., L_2 has an orthogonal basis of eigenfunctions). Bellow and Losert (1985) showed that the sequence of Fourier-Stieltjes coefficients $\{\hat{v}(k)\}_{k \geq 0}$ of a complex Borel measure on $\mathbb{T}$ is a bounded Besicovitch sequence.

By Lin et al. (1999), *Besicovitch sequences are good modulating sequences for every weakly almost periodic operator on a complex Banach space* (even if they are not p -Besicovitch for any $p > 1$), since they are limits in W_1 of bounded Hartman sequences.

Resolvent Conditions and Growth of Powers

The theorems of Koopman and von Neumann (1932), Dunford (1943b), and Katznelson and Tzafriri (1986) show that certain properties of the peripheral spectrum $\sigma(T) \cap \mathbb{T}$ of a power-bounded operator T imply certain ergodic theorems.

The Ritt condition. Ritt (1953) introduced the following spectral condition for T on a complex Banach space with spectral radius $r(T) \leq 1$:

$$\|R(\lambda, T)\| \leq \frac{C}{|\lambda - 1|} \quad \text{for } |\lambda| > 1. \quad (11)$$

Ritt proved that (11) implies $n^{-1}\|T^n\| \rightarrow 0$; the question of whether it implies power-boundedness was studied for many years, until Lyubich (1999) and Nagy and Zemánek (1999) proved (independently) the following.

Theorem 66 *An operator T on a complex Banach space satisfies (11) if and only if $\sup_n \{\|T^n\| + n\|T^n(I - T)\|\} < \infty$.*

Operators satisfying (11) are now called *Ritt operators*; it was also shown by Nagy and Zemánek (1999) that the spectrum of a Ritt operator is contained in a *Stolz region* (the closed convex hull of 1 and a disk of radius $r < 1$). By Theorem 66, *Ritt operators on reflexive Banach spaces are stable* (see Lemma 41). A unitary $T \neq I$ is never Ritt. El-Fallah and Ransford (2002) proved that (11) implies $\|T^n\| \leq C^2$, which for large C was improved by Bakaev (2003) to $\|T^n\| \leq AC \log(AC)$ for some $A > 0$. Lyubich (2001) proved the existence of a Ritt operator $T \neq I$ on L_p with spectrum $\{1\}$. Kalton et al. (2004) proved:

Theorem 67 *Let T on a complex Banach space satisfy $\limsup_{n \rightarrow \infty} n\|T^n - T^{n+1}\| < e^{-1}$; then T is power-bounded, hence a Ritt operator.*

Moreover, they showed that with equality to e^{-1} the result fails in any infinite-dimensional Banach space. Malinen et al. (2009) proved that *if $\|R(\lambda, T)\| \leq C/(\lambda - 1)$ for $1 < \lambda < 1 + \varepsilon$ and $\sup_{n \geq 1} n\|T^n - T^{n+1}\| < \infty$, then T is power-bounded, hence a Ritt operator.* Chen and Shaw (2009) proved that an Abel bounded T satisfying $\sup n\|T^n(I - T)\| < \infty$ is power-bounded, hence Ritt.

Le Merdy and Xu (2012) proved the following:

Theorem 68 *Let $1 < p < \infty$ and let T be a positive contraction of $L_p(\Omega, \mu)$ of a σ -finite measure space. If T is Ritt, then $T^n f$ converges a.e. for every $f \in L_p(\Omega, \mu)$.*

Examples for Theorem 68 were given in Cohen et al. (2014).

Dungey (2011) proved that for every $0 < \alpha < 1$, and every power-bounded T , the operator $S = I - (I - T)^\alpha = \sum_{j=1}^\infty a_j^{(\alpha)} T^j$ (see the section *Rates of convergence*), which is a convex combination of the powers of T , is a Ritt operator. Gomilko and Tomilov (2018) proved that *if T is Ritt, then any convex power series $\sum_{n=0}^\infty c_n T^n$, with $c_n \geq 0$ and $\sum_{n \geq 0} c_n = 1$, is Ritt.*

The Kreiss condition. Kreiss (1962) presented the following resolvent condition (*Kreiss*

resolvent condition) for T on a complex Banach space with $r(T) \leq 1$:

$$\|R(\lambda, T)\| \leq \frac{C}{|\lambda| - 1} \quad \text{for } |\lambda| > 1. \quad (12)$$

Ritt's condition (11) clearly implies (12). Operators satisfying (12) are often called *Kreiss bounded operators*. Power-bounded operators satisfy (12). Kreiss proved that in finite-dimensional spaces (12) implies power-boundedness. Lubich and Nevanlinna (1991) proved that (12) implies $\|T^n\| = O(n)$; by Shields (1978) or Nevanlinna (2001), this is the best estimate. However, Nevanlinna (2001) showed that if T satisfies (12) and its peripheral spectrum $\sigma(T) \cap \mathbb{T}$ has arc-length (Lebesgue) measure zero, then $\|T^n\| = o(n)$. Independently, Cohen et al. (2020) and Bonilla and Müller (2021) proved that in Hilbert spaces, (12) implies $\|T^n\| = O(n/\log n)$. Cuny (2020) proved that a *Kreiss bounded* T on $L_p(\Omega, \mu)$, $1 < p < \infty$, satisfies $\|T^n\| = O(n/\sqrt{\log n})$; he obtained the estimate $\|T^n\| = O(n/(\log n)^{1/s})$, for some $s \geq 2$, when T is Kreiss bounded, defined on a space in a subclass of the spaces which are uniformly convex in an equivalent norm (i.e., UMD spaces, see Cuny (2020)).

Strikwerda and Wade (1997) proved that (12) is equivalent to $\sup_n \sup_{|\gamma|=1} \|C_n^{(2)}(\gamma T)\| < \infty$, where $C_n^{(2)}$ is the Cesàro average of order 2. This characterization was extended by Aleman and Siciu (2016), who proved that T satisfies (12) if and only if for some integer $r \geq 2$ we have $\sup_n \sup_{|\gamma|=1} \|C_n^{(r)}(\gamma T)\| < \infty$. Gomilko and Zemánek (2008) proved that T satisfies (12) if and only if T^m satisfies (12) for some (all) $m > 1$. Siciu and Zemánek (2013) proved that if T satisfies (12), then $C_n^{(2)}(T)x$ converges strongly for $x \in \mathcal{F}(T) \oplus (I - T)\bar{X}$; if in addition $\sigma(T) \cap \mathbb{T} = \{1\}$, then $\|M_{n+1}(T) - M_n(T)\| \rightarrow 0$. Abadias and Bonilla (2019) proved that if T satisfies (12), then for every $\alpha \geq 2$ we have $\|C_{n+1}^{(\alpha)}(T) - C_n^{(\alpha)}(T)\| \rightarrow 0$ (with no assumption on the spectrum). A special case of Siciu

(2016) is that when $\sup_n \|C_n^{(2)}(T)\| < \infty$, then $\|R(\lambda, T)\| \leq C|\lambda - 1|^2/|\lambda - 1|^3$, $|\lambda| > 1$.

Earlier, Kreiss had given a resolvent condition for the generator of a C_0 -semigroup, inspired by the Hille-Yosida theorem, which in finite-dimensional spaces yields boundedness of the semigroup; however, in contrast to the discrete time, Eisner and Zwart (2006) constructed a C_0 -semigroup with *exponential* growth whose generator satisfies Kreiss' condition.

Van Castren (1980) proved that if T is power-bounded invertible on a complex Hilbert space with $\sigma(T) \subset \mathbb{T}$, and T^{-1} satisfies (12) (which is equivalent to condition (ii) in van Casteren's theorem), then also T^{-1} is power-bounded. This extended results of Gokhberg and Krein (1967) and of Stampfli (1972).

McCarthy (1971) gave an example of T invertible on $\ell^2(\mathbb{Z})$ which satisfies the stronger condition (*strong Kreiss resolvent condition*, sometimes called *iterated Kreiss condition*):

$$\|R^k(\lambda, T)\| \leq \frac{C}{(|\lambda| - 1)^k} \quad (13)$$

whenever $|\lambda| > 1$, $k = 1, 2, \dots$

but is not power-bounded; in the example also T^{-1} satisfies (13). By McCarthy (1971), condition (13) implies that $\|T^n\| = O(\sqrt{n})$. This estimate was proved also by Lubich and Nevanlinna (1991), who showed it is the best possible in general Banach spaces. Cohen et al. (2020) showed that in Hilbert space (13) implies $\|T^n\| = O((\log n)^\kappa)$ for some $\kappa > 0$ (which depends on T). Lyubich (2010) obtained a family of examples in $L^p[0, 1]$ satisfying (12) but not (13). Nevanlinna (1997, 2001) proved that T satisfies (13) if and only if for some M we have

$$\|e^{zT}\| \leq M e^{|z|} \quad \forall z \in \mathbb{C}. \quad (14)$$

Gomilko and Zemánek (2013) proved that T satisfies (13) if and only if T^m satisfies it for some (all) integers $m > 1$.

Montes-Rodríguez et al. (2005) defined the *uniform Kreiss resolvent condition* by

$$\sup_{n \geq 1} \left\| \sum_{k=0}^n \frac{T^k}{\lambda^{k+1}} \right\| \leq \frac{C}{|\lambda| - 1} \quad \text{whenever} \quad |\lambda| > 1. \quad (15)$$

They showed that (15) does not imply (13) and proved that (15) holds if and only if there exists $C > 0$ such that

$$\sup_n \left\| \frac{1}{n} \sum_{k=1}^n (\gamma T)^k \right\| \leq C \quad \forall |\gamma| = 1. \quad (16)$$

The proof that (15) implies (12) is immediate. Strikwerda and Wade (1997) showed that (12) does not imply (16). Gomilko and Zemánek (2008) proved that (13) implies (15), hence (16); together with the McCarthy-Lubich-Nevanlinna estimate $\|T^n\| = O(\sqrt{n})$ under (13), Proposition 9 yields:

Theorem 69 *Let T on a reflexive Banach space satisfy the strong Kreiss resolvent condition (13). Then γT is mean ergodic for every $\gamma \in \mathbb{T}$.*

If T is power-bounded, then (13) holds (in an equivalent norm, T is a contraction and $C = 1$ in (12)). Bermúdez et al. (2020) proved

Theorem 70 *Let T on a Hilbert space satisfy the uniform Kreiss resolvent condition (15). Then $\|T^n\| = o(n)$. Hence γT is mean ergodic for every $\gamma \in \mathbb{T}$.*

Theorem 70 was extended by Cuny (2020) to the reflexive L_p spaces (in fact, to the above mentioned UMD spaces). Bonilla and Müller (2021) obtained examples of mean ergodic operators on a complex Hilbert space which do not satisfy the Kreiss resolvent condition (12). Bermúdez et al. (2020) used a strict strengthening of (16) for the following.

Theorem 71 *Let T on a Banach space satisfy $\sup_n n^{-1} \sum_{k=0}^{n-1} \|T^k x\| \leq C \|x\|$ for some $C > 0$. Then $\|T^n\| = o(n)$. When X is reflexive, γT is mean ergodic for every $\gamma \in \mathbb{T}$.*

Cohen et al. (2020) showed that the condition of Theorem 71 (called *absolute Cesàro*

boundedness) and the strong Kreiss resolvent condition (13) are independent. They proved also that if T is absolutely Cesàro bounded, then $\|T^n\| = O(n^{1-\varepsilon})$ for some $\varepsilon > 0$ (which depends on T). Cuny (2020) proved that *when the Banach space X has type $1 < p \leq 2$ (e.g., L_r spaces, $1 < r < \infty$, and more generally spaces which are uniformly convex in an equivalent norm), an absolutely Cesàro bounded T satisfies $\|T^n\| = O(n^{1/p})$.*

Cohen et al. (2020) proved that a positive Cesàro bounded operator on a complex Banach lattice satisfies the uniform Kreiss resolvent condition.

Continuous Time (C_0 -semigroups)

Birkhoff and von Neumann were motivated by a problem in statistical mechanics and therefore proved their ergodic theorems in continuous time. Von Neumann (1932) used in his proof the spectral theorem for unitary representations of $\mathbb{R}$.

A C_0 -semigroup (or *one-parameter semigroup*, *strongly continuous semigroup*) is a family $\{T(t)\}_{t \geq 0}$ of bounded linear operators on a Banach space X such that $T(0) = I$, $T(t+s) = T(t)T(s)$ for $t, s \geq 0$, and $T(t)x$ is continuous on $[0, \infty)$ for every $x \in X$. The limit $\omega_0 := \lim_{t \rightarrow \infty} \log \|T(t)\|/t < \infty$ (exists by subadditivity) is the *type* (or *growth bound*) of the semigroup, and for any $\delta > \omega_0$ there exists M_δ such that $\|T(t)\| \leq M_\delta e^{\delta t}$ for $t \geq 0$. The *generator* (*infinitesimal generator*) of a C_0 -semigroup $\{T(t)\}_{t \geq 0}$ is the operator $Ax := \lim_{t \rightarrow 0+} t^{-1}(T(t)x - x)$ with domain $\mathcal{D}(A) := \{x \in X : Ax \text{ exists}\}$. For properties of the generator, we refer the reader to Section VIII.1 of the book by Dunford and Schwartz (1958):

- (i) The domain of the generator is dense in X , and A is a closed operator.
- (ii) For $x \in \mathcal{D}(A)$ and $t > 0$, $T(t)x \in \mathcal{D}(A)$ and $\frac{d}{dt}T(t)x = AT(t)x = T(t)Ax$.
- (iii) The generator is a bounded operator if and only if the semigroup is continuous in the uniform (operator norm) topology.

- (iv) If $\operatorname{Re} \lambda > \omega_0$, then $\lambda \in \rho(A)$, and $R(\lambda, A)x = \int_0^\infty e^{-\lambda t} T(t)x dt$ for every $x \in X$.

Hille (1952) introduced the following abstract generalization (*the abstract Cauchy problem*) of the (homogenous) initial value problem of partial differential equations. Let A be a linear operator (not necessarily bounded) mapping a subspace $\mathcal{D}$ (not necessarily closed) of a complex Banach space X into X , and consider the problem of finding a differentiable function $y(t)$, defined for $t > 0$ with values in $\mathcal{D}$, such that $\frac{d}{dt}y(t) = Ay(t)$ for $t > 0$, and $y(t) \rightarrow y_0$ as $t \rightarrow 0^+$ for a given y_0 .

When $\mathcal{D} = X$ and A is bounded, the operator e^{tA} is a well-defined bounded operator, and $y(t) := e^{tA}y_0$ solves the abstract Cauchy problem. Hille noted that if A is the generator of a C_0 -semigroup $T(t)$, then $y(t) := T(t)y_0$ solves abstract Cauchy problem when $y_0 \in \mathcal{D}(A)$.

Ergodic theorems for a C_0 -semigroup $\{T(t)\}$ deal with the asymptotic behavior of $T(t)x$ as $t \rightarrow \infty$, in different topologies and different modes of convergence. These yield information on the asymptotic behavior of the solutions of the abstract Cauchy problem when A is the generator of a C_0 -semigroup; from this perspective, it is important to use assumptions only on the generator of the semigroup, which is the given datum of the problem. The book of van Neerven (1996) studies the different spectral properties of the generator (which is usually only an unbounded closed operator), and their connections to the asymptotic behavior of the semigroup. The paper of Rozendaal and Veraar (2018), which uses growth rates of the resolvent of the generator for obtaining growth rates of the semigroup, contains references to papers which appeared after publication of van Neerven's book.

A nonspectral approach is in the book of Emel'yanov (2007), which emphasizes C_0 -semigroups of positive operators on Banach lattices, in particular C_0 -semigroups of positive operators on L_1 spaces.

Many results on the asymptotic behavior of C_0 -semigroups have analogues for discrete time, and sometimes the continuous time version was proved first (see Theorems 25, 38, and 46).

A C_0 -semigroup $\{T(t)\}_{t \geq 0}$ is called *mean (uniformly) ergodic* if the averages $M_s x := \frac{1}{s} \int_0^s T(t)x dt$ converge in the strong (uniform) operator topology, as $s \rightarrow \infty$. The limit is a projection on the common fixed points of $\{T(t)\}_{t \geq 0}$.

Theorem 72 *A bounded C_0 -semigroup on a reflexive Banach space is mean ergodic.*

This continuous version of Lorch's Theorem (Corollary 8) follows from the abstract ergodic theorems of Eberlein (1949). Hille (see Section 18.6 of Hille and Phillips (1957), Section V.4 of Engel and Nagel (2000)) proved the following.

Theorem 73 *Let $\{T(t)\}_{t \geq 0}$ be a mean ergodic C_0 -semigroup on X , with generator A . Then $X = \ker(A) \oplus \overline{\operatorname{Range}(A)}$, and $E_x := \lim_{s \rightarrow \infty} M_s x$ is the projection on $\ker(A)$ corresponding to this decomposition.*

Lin (1974b) proved the continuous analogue of Theorem 23.

Theorem 74 *A C_0 -semigroup with generator A is uniformly ergodic if and only if $\lim_{t \rightarrow \infty} t^{-1} \|T(t)\| = 0$, and the range of A is closed.*

Krengel and Lin (1984) proved the continuous analogue of Browder's Theorem 14.

Theorem 75 *Let $\{T(t)\}_{t \geq 0}$ be a bounded C_0 -semigroup on a reflexive Banach space, with generator A . Then y is in the range of A if and only if $\sup_{s > 0} \left\| \int_0^s T(t)x dt \right\| < \infty$.*

Eisner (2010) emphasizes the similarities of the discrete and continuous cases and proves in parallel results for discrete time and their analogues for continuous time. However, not all results in the discrete case have analogues in continuous time, e.g., Eisner and Zwart (2006), and some results for semigroups have no analogue in discrete time, e.g., Gerlach (2013), where it is shown that a convergence result for certain C_0 -semigroups fails in discrete time under the analogous assumptions, due to some periodic behavior.

Bibliography

The references below are the primary sources cited in this article. Most of them contain additional relevant results and references. Many of the listed authors have additional related papers, not cited in the present article

All books on ergodic theory have a discussion of the mean ergodic theorem. We list below only books not cited in the article, which have chapters containing more operator theoretic, including asymptotic behavior of C_0 -semigroups

References

- Aaronson J, Weiss B (2000) Remarks on the tightness of cocycles. *Colloq Math* 84/85(part 2):363–376
- Abadias L, Bonilla A (2019) Growth orders and ergodicity for absolutely Cesàro bounded operators. *Linear Algebra Appl* 561:253–267
- Akcoglu MA (1975) A pointwise ergodic theorem in L_p -Spaces. *Canadian J Math* 27:1075–1082
- Akcoglu M, Sucheston L (1972) On operator convergence in Hilbert space and in Lebesgue space. *Period Math Hung* 2:235–244
- Akcoglu M, Sucheston L (1975) Weak convergence of positive contractions implies strong convergence of averages. *Z Wahrscheinlichkeitstheorie Verw Gebiete* 32:139–145
- Akcoglu MA, Huneke JP, Rost H (1974) A counter example to the Blum Hanson theorem in general spaces. *Pac J Math* 50:305–308
- Aleman A, Suciú L (2016) On ergodic operator means in Banach spaces. *Integral Eq Operator Theory* 85: 259–287
- Allan GR (1997) Power-bounded elements and radical Banach algebras. In: *Linear operators*, vol 38. Banach Center Publications, Polish Academy of Sciences, Institute of Mathematics, Warsaw, pp 9–16
- Allan GR, Ransford T (1989) Power-dominated elements in a Banach algebra. *Stud Math* 94:63–79
- Allan GR, O’Farrell AG, Ransford TJ (1987) A Tauberian theorem arising in operator theory. *Bull Lond Math Soc* 19:537–545
- Anosov D (1973) The additive functional homology equation that is connected with an ergodic rotation of the circle. *Izv. Akad Nauk SSSR Ser Mat* 37:1257–1271. (Russian; English transl. in *Math. USSR Izvestya* 7 (1973))
- Arendt W, Batty C (1988) Tauberian theorems and stability of one-parameter semigroups. *Trans Am Math Soc* 306: 837–852
- Assani I (1986) Sur les opérateurs à puissances bornées et le théorème ergodique ponctuel dans $L^p[0, 1]$ $1 < p < +\infty$. *Can J Math* 38:937–946
- Assani I (2003) *Wiener Wintner ergodic theorems*. World Scientific, River Edge
- Avigad J, Rute J (2015) Oscillation and the mean ergodic theorem for uniformly convex Banach spaces. *Ergodic Theory Dynam Syst* 35:1009–1027
- Azizov AN, Chilin VI (2017) The Blum-Hanson ergodic theorem in Banach lattices of sequences (Russian). *Vladikavkaz Mat Zh* 19:3–10
- Badea C, Müller V (2009) On weak orbits of operators. *Topology Appl* 156:1381–1385
- Badea C, Seifert D (2017) Quantified asymptotic behaviour of Banach space operators and applications to iterative projection methods. *Pure Appl Funct Anal* 2: 585–598
- Badiozzaman AJ, Thorpe B (1992) A uniform ergodic theorem and summability. *Bull Lond Math Soc* 24: 351–360
- Bakaev NY (2003) Constant size control in stability estimates under some resolvent conditions. *Numerical Methods Program* 4:348–357
- Banach S (1932) *Théorie des Opérations Linéaires*. Monografie Matematyczne, Warsaw
- Bartoszek W (1988) Asymptotic periodicity of the iterates of positive contractions on Banach lattices. *Stud Math* 91:179–188
- Baskakov AG (2015) Harmonic and spectral analysis of power bounded operators and bounded semigroups of operators on a Banach space (Russian). *Mat Zametki* 97:174–190; English translation: *Math Notes* 97 (2015), 164–178
- Batty CJK (1994) Asymptotic behaviour of semigroups of operators. In: *Functional analysis and operator theory*, Banach Center Publication 30. Polish Academy of Sciences Institute of Mathematics, Warsaw, pp 35–52
- Batty CJK, Vū QP (1990) Stability of individual elements under one-parameter semigroups. *Trans Am Math Soc* 322:805–818
- Batty CJK, Seifert D (2022) Some developments around the Katznelson-Tzafriri theorem. *Acta Sci Math (Szeged)* 88:53–84
- Becker ME (2005) A condition equivalent to uniform ergodicity. *Stud Math* 167:215–218
- Becker ME (2011) Ergodic properties of linear operators. *Rev Un Mat Argentina* 52:23–25
- Bellow A (1975) An L_p -inequality with application to ergodic theory. *Houst J Math* 1:153–159
- Bellow A, Losert V (1985) The weighted pointwise ergodic theorem and the individual ergodic theorem along subsequences. *Trans Am Math Soc* 288:307–345
- Berend D (1992) A note on the L^p analogue of the “zero-two” law. *Proc Am Math Soc* 114:95–97
- Berend D, Bergelson V (1986) Mixing sequences in Hilbert spaces. *Proc Am Math Soc* 98:239–246
- Berend D, Lin M, Rosenblatt J, Tempelman A (2002) Modulated and subsequential ergodic theorems in Hilbert and Banach spaces. *Ergodic Theory Dyn Sys* 22:1653–1665
- Bergelson V (1987) Weakly mixing PET. *Ergodic Theory Dyn Syst* 7:337–349
- Berkson E, Gillespie TA (1987) Stečkin’s theorem, transference, and spectral decompositions. *J Funct Anal* 70: 140–170
- Bermúdez T, González M, Mbekhta M (2000) Operators with an ergodic power. *Stud Math* 141:201–208

- Bermúdez T, Bonilla A, Müller V, Peris A (2020) Cesàro bounded operators in Banach spaces. *J Anal Math* 140: 187–206
- Besicovitch AS (1954) Almost periodic functions. Dover Publications, New York
- Birkhoff GD (1931) Proof of the ergodic theorem. *Proc Natl Acad Sci U S A* 17:656–660
- Birkhoff G (1939) The mean ergodic theorem. *Duke Math J* 5:19–20
- Blum JR, Eisenberg B (1974) Generalized summing sequences and the mean ergodic theorem. *Proc Am Math Soc* 42:423–429
- Blum JR, Hanson DL (1960) On the mean ergodic theorem for subsequences. *Bull Am Math Soc* 66:308–311
- Bonilla A, Müller V (2021) Kreiss bounded and uniformly Kreiss bounded operators. *Rev Mat Complut* 34: 469–487
- Boshernitzan M, Wierdl M (1996) Ergodic theorems along sequences and Hardy fields. *Proc Natl Acad Sci U S A* 93:8205–8207
- Bourgain J (1988a) An approach to pointwise ergodic theorems. In: *Geometric aspects of functional analysis* (1986/87), Lecture notes in Mathematics 1317. Springer, Berlin, pp 204–223
- Bourgain J (1988b) Temps de retour pour les systèmes dynamiques. *C R Acad Sci* 306(12):483–485
- Bourgain J (1989) Pointwise ergodic theorems for arithmetic sets. *Inst Hautes Études Sci Publ Math* 69:5–45
- Bourgain J, Furstenberg H, Katznelson Y, Ornstein DS (1989) On return-time sequences, Appendix to Bourgain (1989). *Inst Hautes Études Sci Publ Math* No 69:42–45
- Browder F (1958) On the iteration of transformations in noncompact minimal dynamical systems. *Proc Am Math Soc* 9:773–780
- Brunel A, Keane M (1969) Ergodic theorems for operator sequences. *Z Wahrscheinlichkeitstheorie und Verw Gebiete* 12:231–240
- Brunel A, Revuz D (1974) Quelques applications probabilistes de la quasi-compacité. *Ann Inst H Poincaré Sect B (NS)* 10:301–337
- Burlando L (1997) A generalization of the uniform ergodic theorem to poles of arbitrary order. *Stud Math* 122:75–98
- Butzer PL, Westphal U (1971) The mean ergodic theorem and saturation. *Indiana Univ Math J* 20:1163–1174
- Butzer PL, Westphal U (1972) Ein Operatorenkalkül für das approximationstheoretische Verhalten des Ergodensatzes im Mittel. In: *Linear operators and approximation*, International series of numerical mathematics, vol 20. Birkhäuser, Basel, pp 102–114
- Campbell JT (1986) Spectral analysis of the ergodic Hilbert transform. *Indiana Univ Math J* 35:379–390
- Chacon RV (1969) Weakly mixing transformations which are not strongly mixing. *Proc Am Math Soc* 22:559–562
- Chen J-C, Shaw S-Y (2009) Growth order and stability of discrete semigroups. *Nonlinear Anal* 71(12):e2879–e2882
- Chill R, Tomilov Y (2007) Stability of operator semigroups: ideas and results. In: *Perspectives in operator theory*, Banach Center Publications, vol 75. Polish Academy of Science, Warsaw, pp 71–109
- Cohen LW (1940) On the mean ergodic theorem. *Ann Math* 41(2):505–509
- Cohen G, Lin M (2016) Remarks on rates of convergence of powers of contractions. *J Math Anal Appl* 436: 1196–1213
- Cohen G, Cuny C, Lin M (2010) The one-sided ergodic Hilbert transform in Banach spaces. *Stud Math* 196: 251–263
- Cohen G, Cuny C, Lin M (2014) Almost everywhere convergence of powers of some positive L_p contractions. *J Math Anal Appl* 420:1129–1153
- Cohen G, Cuny C, Eisner T, Lin M (2020) Resolvent conditions and growth of powers of operators. *J Math Anal Appl* 487:124035, 24 p
- Çömez D, Lin M, Olsen J (1998) Weighted ergodic theorems for mean ergodic L_1 -contractions. *Trans Am Math Soc* 350:101–117
- Conze J-P (1973) Convergence des moyennes ergodiques pour des sous-suites. In: *Contributions au calcul des probabilités*, Bulletin Society Mathematics, France, Mém No. 35, Society Mathematics, France, pp 7–15
- Cotlar M (1955) A unified theory of Hilbert transforms and ergodic theorems. *Rev Mat Cuyana* 1:105–167
- Cuny C (2009) On the a.s. convergence of the one-sided ergodic Hilbert transform. *Ergodic Theory Dyn Syst* 29:1781–1788
- Cuny C (2010) Norm convergence of some power series of operators in L^p with applications in ergodic theory. *Stud Math* 200:1–29
- Cuny C (2020) Resolvent conditions and growth of powers of operators on L^p spaces. *Pure Appl Funct Anal* 5: 1025–1038
- Cuny C, Weber M (2018) On Nörlund summation and ergodic theory, with applications to power series of Hilbert contractions. *Bull Pol Acad Sci Math* 66:69–85
- Datko R (1970) Extending a theorem of A. M. Liapunov to Hilbert space. *J Math Anal Appl* 32:610–616
- deLaubenfels R, Vũ Q-P (1996) The discrete Hille-Yosida space and the asymptotic behaviour of individual orbits of linear operators. *J Funct Anal* 142:539–548
- deLeeuw K, Glicksberg I (1961) Applications of almost periodic compactifications. *Acta Math* 105:63–97
- Demeter C (2012) Improved range in the return times theorem. *Can Math Bull* 55:708–722
- Demeter C, Lacey MT, Tao T, Thiele C (2008) Breaking the duality in the return times theorem. *Duke Math J* 143:281–355
- Dénial Y (1989) On the a.s. Cesàro- α convergence for stationary or orthogonal random variables. *J Theor Probab* 2:475–485
- Derriennic Y (1976) Lois “zéro ou deux” pour les processus de Markov. Applications aux marches aléatoires. *Ann Inst H Poincaré Sect B (NS)* 12:111–129
- Derriennic Y (2000) On the mean ergodic theorem for Cesàro bounded operators. *Colloq Math* 84/85(part 2):443–455

- Derriennic Y, Krengel U (1981) Subadditive mean ergodic theorems. *Ergodic Theory Dyn Syst* 1:33–48
- Derriennic Y, Lin M (2001) Fractional Poisson equations and ergodic theorems for fractional coboundaries. *Israel J Math* 123:93–130
- Devys O (2012) Localisation spectrale à l'aide des polynômes de Faber et équation de cobord. PhD thesis, Lille. <https://pepite-depot.univ-lille.fr/LIBRE/EDSPI/2012/50376-2012-Devys.pdf>
- Doebelin W (1937) Sur les propriétés asymptotiques de mouvements régis par certains types de chaînes simples. *Bull Soc Math Roumaine* 39–1:57–115; 39–2 (1938), 3–61
- Drissi D, Zemánek J (2000) Gelfand-Hille theorems for Cesàro means. *Quaest Math* 23:375–381
- Dunford N (1943a) Spectral theory, I. Convergence to projections. *Trans Am Math Soc* 54:185–217
- Dunford N (1943b) Spectral theory. *Bull Am Math Soc* 49: 637–651
- Dunford N, Schwartz JT (1956) Convergence almost everywhere of operator averages. *J Rational Mech Anal* 5:129–178
- Dungey N (2008) On time regularity and related conditions for power-bounded operators. *Proc Lond Math Soc* 97(3):97–116
- Dungey N (2011) Subordinated discrete semigroups of operators. *Trans Am Math Soc* 363:1721–1741
- Eberlein WF (1949) Abstract ergodic theorems and weak almost periodic functions. *Trans Am Math Soc* 67: 217–240
- Eberlein WF (1984) On retrogression in mean ergodic theory. *J Approx Theory* 42:293–298
- Ed-dari E (2003) On the (C, α) uniform ergodic theorem. *Stud Math* 156:3–13
- Ed-dari E (2004) On the (C, α) Cesàro bounded operators. *Stud Math* 161:163–175
- Eisner T (2010) Stability of operators and operator semigroups, *Operator theory: advances and applications*, vol 209. Birkhäuser, Basel
- Eisner T (2013) Linear sequences and weighted ergodic theorems. *Abstr Appl Anal*:815726, 5 p
- Eisner T, Lin M (2018) On modulated ergodic theorems. *J Nonlinear Var Anal* 2:131–154
- Eisner T, Müller V (2021) Power bounded operators and the mean ergodic theorem for subsequences. *J Math Anal Appl* 493:124523, 25 p
- Eisner T, Serény A (2008) Category theorems for stable operators on Hilbert spaces. *Acta Sci Math (Szeged)* 74:259–270
- Eisner T, Zwart H (2006) Continuous-time Kreiss resolvent condition on infinite-dimensional spaces. *Math Comput* 75(256):1971–1985
- El Abdalaoui EH, El Machkouri M, Nogueira A (2010) A criterion of weak mixing property. *Séminaires et Congrès* 20:105–111
- El-Fallah O, Ransford T (2002) Extremal growth of powers of operators satisfying resolvent conditions of Kreiss-Ritt type. *J Funct Anal* 196:135–154
- Emel'yanov EY (1997) Banach lattices on which every power-bounded operator is mean ergodic. *Positivity* 1: 291–295
- Émilion R (1985) Mean-bounded operators and mean ergodic theorems. *J Funct Anal* 61:114
- Esterle J (1983) Quasimultipliers, representations of H^∞ , and the closed ideal problem for commutative Banach algebras. In: *Radical Banach algebras and automatic continuity*, Lecture Notes in Mathematics 975. Springer, Berlin, pp 66–162
- Esterle J, Strouse E, Zouakia F (1990) Theorems of Katznelson-Tzafriri type for contractions. *J Funct Anal* 94:273–287
- Esterle J, Strouse E, Zouakia F (1992) Stabilité asymptotique de certains semi-groupes d'opérateurs et idéaux primaires de $L^1(R^+)$. *J Operator Theory* 28:203–227
- Fan A-H (2019) Weighted Birkhoff ergodic theorem with oscillating weights. *Ergodic Theory Dyn Syst* 39: 1275–1289
- Farkas B, Kreidler H (2021) Relative compactness of orbits and geometry of Banach spaces. *J Math Anal Appl* 495:124660
- Foguel S (1963) Powers of a contraction in Hilbert space. *Pac J Math* 13:551–562
- Foguel S (1976) More on the “zero-two” law. *Proc Am Math Soc* 61:262–264
- Foguel S, Weiss B (1973) On convex power series of a conservative Markov operator. *Proc Am Math Soc* 38: 325–330
- Foiaş C (1964) Sur les mesures spectrales qui interviennent dans la théorie ergodique. *J Math Mech* 13:639–658
- Fonf VP, Lin M, Rubinov A (1996) On the uniform ergodic theorem in Banach spaces that do not contain duals. *Stud Math* 121:67–85
- Fonf VP, Lin M, Wojtaszczyk P (2001) Ergodic characterizations of reflexivity of Banach spaces. *J Funct Anal* 187:146–162
- Fonf VP, Lin M, Wojtaszczyk P (2010) A non-reflexive Banach space with all contractions mean ergodic. *Israel J Math* 179:479–491
- Fonf VP, Lin M, Wojtaszczyk P (2011) Poisson's equation and characterizations of reflexivity of Banach spaces. *Colloq Math* 124:225–235
- Fong H, Sucheston L (1974) On a mixing property of operators in L_p spaces. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 28:165–171
- Furstenberg H (1981) Recurrence in ergodic theory and combinatorial number theory. Princeton University Press, Princeton
- Gaposhkin VF (1998) On the rate of decrease of the probabilities of ϵ -deviations for means of stationary processes. *Mat Zametki* 64(3):366–372; English translation in *Math Notes* 64 (1998), 316–321
- Gelfand I (1941) Zur Theorie der Charaktere der Abelschen topologischen Gruppen. *Mat Sbornik* 9:49–50
- Gerlach M (2013) On the peripheral point spectrum and the asymptotic behavior of irreducible semigroups of Harris operators. *Positivity* 17:875–898

- Gerlach M, Glück J (2019) Mean ergodicity vs weak almost periodicity. *Stud Math* 248:45–56
- Glück J (2015) On weak decay rates and uniform stability of bounded linear operators. *Arch Math* 104:347–356
- Gokhberg I, Krein MG (1967) Description of contraction operators which are similar to unitary operators, *Functional. Anal Appl* 1:33–52
- Goldstein JA (1993) Extremal properties of contraction semigroups on Hilbert and Banach spaces. *Bull Lond Math Soc* 25:369–376
- Goldstein JA, Nagy B (1995) An extremal property of contraction semigroups in Banach spaces. *Ill J Math* 39:441–449
- Gomilko A, Tomilov Y (2018) On discrete subordination of power bounded and Ritt operators. *Indiana Univ Math J* 67:781–829
- Gomilko A, Zemánek J (2008) On the uniform Kreiss resolvent condition. *Funct Anal Appl* 42:230–233
- Gomilko A, Zemánek J (2013) On the strong Kreiss resolvent condition. *Complex Anal Oper Theory* 7:421–435
- Gomilko A, Haase M, Tomilov Y (2011) On rates in mean ergodic theorems. *Math Res Lett* 18:201–213
- Gottschalk W, Hedlund GA (1955) Topological dynamics, American Mathematical Society Colloquium Publications, vol 36. American Mathematical Society, Providence
- Grabiner S, Zemánek J (2002) Ascent, descent, and ergodic properties of linear operators. *J Operator Theory* 48: 69–81
- Grivaux S (2019) The Blum-Hanson property. *Concr Oper* 6:92–105
- Grobler JJ, Huijsmans CB (1995) Doubly Abel bounded operators with single spectrum. *Quaest Math* 18: 397–406
- Groh U (1983) On the peripheral spectrum of uniformly ergodic positive operators on C^* -algebras. *J Operator Theory* 10:31–37
- Haase M, Tomilov Y (2010) Domain characterizations of certain functions of power-bounded operators. *Stud Math* 196:265–288
- Halmos PR (1944) In general a measure preserving transformation is mixing. *Ann Math* 45(2):786–792
- Halmos PR (1949a) Measurable transformations. *Bull Am Math Soc* 55:1015–1034
- Halmos PR (1949b) A nonhomogeneous ergodic theorem. *Trans Am Math Soc* 66:284–288
- Hanson DL, Pledger G (1969) On the mean ergodic theorem for weighted averages. *Z Wahrscheinlichkeitstheorie und Verw Gebiete* 13:141–149
- Hardy GH, Littlewood JE (1913) Sur la série de Fourier d'une fonction à carré sommable. *C. R. Acad Sci Paris* 156:1307–1309
- Hiai F (1978) Weakly mixing properties of semigroups of linear operators. *Kodai Math J* 1:376–393
- Hille E (1945) Remarks on ergodic theorems. *Trans Am Math Soc* 57:246–269
- Hille E (1952) Une généralisation du problème de Cauchy. *Ann Inst Fourier (Grenoble)* 4:31–48
- Hopf E (1932) Proof of Gibbs' hypothesis on the tendency toward statistical equilibrium. *Proc Nat Acad Sci U S A* 18:33–340
- Hopf E (1937) Ergodentheorie, *Ergebnisse der Mathematik und ihrer Grenzgebiete* 5, No. 2. Julius Springer
- Horowitz S (1968) Some limit theorems for Markov processes. *Israel J Math* 6:107–118
- Horowitz S (1969) Strong ergodic theorems for Markov processes. *Proc Am Math Soc* 23:328–334
- Huang SZ (1995) Stability properties characterizing the spectra of operators on Banach spaces. *J Funct Anal* 132:361–382
- Ionescu Tulcea A (1963) Random series and spectra of measure-preserving transformations. In: *Ergodic theory, Proceedings of the International Symposium, Tulane University, New Orleans*. Academic Press, Boston, pp 273–292
- Irmisch R (1980) Punktweise Ergodensätze für (c, α) -Verfahren, $0 < \alpha < 1$. PhD thesis, Darmstadt
- Jacobs K (1957) Fastperiodizitätseigenschaften allgemeiner Halbgruppen in Banach-Räumen. *Math Z* 67:83–92
- Jamison B (1965) Eigenvalues of modulus 1. *Proc Am Math Soc* 16:375–377
- Jones L (1971) A mean ergodic theorem for weakly mixing operators. *Adv Math* 7:211–216
- Jones L, Kuftinec V (1971) A note on the Blum-Hanson theorem. *Proc Am Math Soc* 30:202–203
- Jones L, Lin M (1976) Ergodic theorems of weak mixing type. *Proc Am Math Soc* 57:50–52
- Jones L, Lin M (1980) Unimodular eigenvalues and weak mixing. *J Funct Anal* 35:42–48
- Jones RL, Ostrovskii IV, Rosenblatt JM (1996) Square functions in ergodic theory. *Ergodic Theory Dyn Syst* 16:267–305
- Jones RL, Kaufman R, Rosenblatt JM, Wierdl M (1998) Oscillation in ergodic theory. *Ergodic Theory Dyn Syst* 18:889–935
- Kachurovskii AG, Podvigin IV (2016) Estimates of the rate of convergence in the von Neumann and Birkhoff ergodic theorems. *Trans Moscow Math Soc*:1–53
- Kachurovskii AG, Sedalishchev VV (2010) On the constants in the estimates for the rate of convergence in von Neumann's ergodic theorem. *Mat Zametki* 87: 756–763, (Russian; English translation in *Math. Notes* 87 (2010), 720–727)
- Kahane J-P (1961) Sur les coefficients de Fourier-Bohr. *Stud Math* 21:103–106
- Kakutani S (1938) Iteration of linear operations in complex Banach spaces. *Proc Imp Acad Tokyo* 14:295–300
- Kakutani S (1950) Ergodic theory. In: *Proceedings of the 1950 International Congress of Mathematicians, Cambridge*, pp 128–142
- Kakutani S, Petersen K (1981) The speed of convergence in the ergodic theorem. *Monatsh Math* 91:11–18
- Kalton N, Montgomery-Smith S, Oleszkiewicz K, Tomilov Y (2004) Power-bounded operators and related norm estimates. *J Lond Math Soc* 70:463–478

- Katznelson Y, Tzafriri L (1986) On power bounded operators. *J Funct Anal* 68:313–328
- Kendall DG, Reuter GEH (1956) Some ergodic theorems for one-parameter semigroups of operators. *Philos Trans R Soc Lond Ser A* 249:151–177
- Kérchy L, van Neerven J (1997) Polynomially bounded operators whose spectrum on the unit circle has measure zero. *Acta Sci Math (Szeged)* 63:551–562
- Kohlenbach U, Leuştean L (2009) A quantitative mean ergodic theorem for uniformly convex Banach spaces. *Ergodic Theory Dyn Syst* 29:1907–1915. Erratum: p 1995
- Koliha JJ (1974) Power convergence and pseudoinverses of operators in Banach spaces. *J Math Anal Appl* 48: 446–469
- Komorník J (1993) Asymptotic periodicity of Markov and related operators. In: *Dynamics reported, Dynamical report expositions dynamics systems (N.S.), 2*, Springer, Berlin, pp 31–68
- Koopman BO, von Neumann J (1932) Dynamical systems of continuous spectra. *Proc Natl Acad Sci USA* 18: 255–263
- Kornfeld I, Kosek W (2003) Positive L^1 operators associated with nonsingular mappings and an example of E. Hille. *Colloq Math* 98:63–77
- Kornfeld I, Lin M (1997) Coboundaries of irreducible Markov operators on $C(K)$. *Israel J Math*. 97:189–202
- Kosek W (2011) Example of a mean ergodic L^1 operator with the linear rate of growth. *Colloq Math* 124(1): 15–22
- Kozitsky Y, Shoikhet D, Zemánek J (2013) Power convergence of Abel averages. *Arch Math* 100:539–549
- Kozma G, Lev N (2011) Exponential Riesz bases, discrepancy of irrational rotations and BMO. *J Fourier Anal Appl* 17:879–898
- Krein MG, Šmulian V (1940) On regularly convex sets in the space conjugate to a Banach space. *Ann Math* 41(2):556–583
- Kreiss H-O (1962) ber die Stabilitätsdefinition für Differenzgleichungen die partielle Differentialgleichungen approximieren. *BIT* 2: 153–181
- Krengel U (1971) On the individual ergodic theorem for subsequences. *Ann Math Statist* 42:1091–1095
- Krengel U (1972) Weakly wandering vectors and weakly independent partitions. *Trans Am Math Soc* 164: 199–226
- Krengel U (1978) On the speed of convergence in the ergodic theorem. *Monatsh Math* 86:3–6
- Krengel U, Lin M (1984) On the range of the generator of a Markovian semigroup. *Math Z* 185:553–565
- Krotkov V, Halperin I (1953) The ergodic theorem for Banach spaces with convex-compactness. *Trans R Soc Canada Sect III* 47:17–20
- Kryloff N, Bogoliouboff N (1937) Sur les propriétés en chaîne. *C R Acad Sci Paris* 204:1386–1388
- Kunszenti-Kovács D (2015) Almost weak polynomial stability of operators. *Houst J Math* 41:901–913
- Kunszenti-Kovács D, Nittka R, Sauter M (2011) On the limits of Cesàro means of polynomial powers. *Math Z* 268:771–776
- Lasota A, Li TY, Yorke JA (1984) Asymptotic periodicity of the iterates of Markov operators. *Trans Am Math Soc* 286:751–764
- Le Merdy C, Xu Q (2012) Strong q -variation inequalities for analytic semigroups. *Ann Inst Fourier* 62: 2069–2097
- Lefèvre P, Matheron É (2016) The Blum-Hanson property for $C(K)$ spaces. *Pac J Math* 282:203–212
- Lefèvre P, Matheron É, Primot A (2016) Smoothness, asymptotic smoothness and the Blum-Hanson property. *Israel J Math* 211:271–309
- Léka Z (2009) A Katznelson-Tzafriri type theorem in Hilbert spaces. *Proc Am Math Soc* 137:3763–3768
- Léka Z (2010) A note on the powers of Cesàro bounded operators. *Czechoslov Math J* 60(135):1091–1100
- Leonov VP (1961) On the dispersion of time means-dependent of a stationary stochastic process. *Teor Veroyatn Primen* 6:93–101, in Russian; English transl in *Theory Probab Appl* 6 (1961), 87–93
- Li Y-C, Satō R, Shaw S-Y (2008) Boundedness and growth orders of means of discrete and continuous semigroups of operators. *Stud Math* 187:1–35
- Lin M (1974a) On the uniform ergodic theorem. *Proc Am Math Soc* 43:337–340
- Lin M (1974b) On the uniform ergodic theorem II. *Proc Am Math Soc* 46:217–225
- Lin M (1978) Quasi-compactness and uniform ergodicity of positive operators. *Israel J Math* 29:309–311
- Lin M (1998) The uniform zero-two law for positive operators in Banach lattices. *Studia Math* 131:149–153
- Lin M, Sine R (1983) Ergodic theory and the functional equation $(I-T)x=y$. *J Operator Theory* 10:153–166
- Lin M, Suciu L (2015) Poisson's equation for mean ergodic operators. In: *Infinite products of operators and their applications*, Contemporary mathematics 636, Proceedings of the American Mathematical Society, Providence, pp 141–148
- Lin M, Weber M (2007) Weighted ergodic theorems and strong laws of large numbers. *Ergodic Theory Dyn Syst* 27:511–543
- Lin M, Olsen J, Tempelman A (1999) On modulated ergodic theorems for Dunford-Schwartz operators. *Ill J Math* 43:542–567
- Lin M, Shoikhet D, Suciu L (2015) Remarks on uniform ergodic theorems. *Acta Sci Math (Szeged)* 81:251–283
- Lorch ER (1939) Means of iterated transformations in reflexive vector spaces. *Bull Am Math Soc* 45:945–947
- Lotz H (1968) Über das Spektrum positiver Operatoren. *Math Z* 108:15–32
- Lotz H (1981) Uniform ergodic theorems for Markov operators on $C(X)$. *Math Z* 178:145–156
- Lotz H (1985) Uniform convergence of operators on L^∞ and similar spaces. *Math Z* 190:207–220
- Lubich C, Nevanlinna O (1991) On resolvent conditions and stability estimates. *BIT* 31:293–313

- Luecke G (1977) Norm convergence of T^n . Canadian J Math 29:1340–1344
- Lyubich Y (1999) Spectral localization, power boundedness and invariant subspaces under Ritt's type condition. Stud Math 134:153–167
- Lyubich Y (2001) The single-point spectrum operators satisfying Ritt's resolvent condition. Stud Math 145:135–142
- Lyubich Y (2010) The power boundedness and resolvent conditions for functions of the classical Volterra operator. Stud Math 196:41–63
- Lyubich YI, Vũ QP (1988) Asymptotic stability of linear differential equations in Banach spaces. Stud Math 88:37–42
- Malinen J, Nevanlinna O, Turunen V, Yuan Z (2007) A lower bound for the difference of powers of linear operators. Acta Math Sin (Engl Ser) 23:745–748
- Malinen J, Nevanlinna O, Yuan Z (2009) On a Tauberian condition for bounded linear operators. Math Proc R Ir Acad 109:101–108
- Mbekhta M, Zemánek J (1993) Sur le théorème ergodique uniforme et le spectre. C R Acad Sci Paris Sr I Math 317(12):1155–1158
- McCarthy C (1971) A strong resolvent condition does not imply power-boundedness. Chalmers Inst Technol Univ Gothenburg preprint 15
- Michal AD, Wyman M (1941) Characterization of complex couple spaces. Ann Math 42(2):247–250
- Millet A (1976) Un théorème ergodique en moyenne. C. R. Acad Sci Paris Sér A-B 283(16):A1103–A1106
- Montes-Rodríguez A, Sánchez-Álvarez J, Zemánek J (2005) Uniform Abel-Kreiss boundedness and the extremal behaviour of the Volterra operator. Proc Lond Math Soc 91(3):761–788
- Mugnolo D (2004) A semigroup analogue of the Fonf-Lin-Wojtaszczyk ergodic characterization of reflexive Banach spaces with a basis. Stud Math 164:243–251
- Müller V, Tomilov Y (2007) Quasimilarity of power bounded operators and Blum-Hanson property. J Funct Anal 246:385–399
- Müller V, Tomilov Y (2010) Weakly wandering vectors and interpolation theorems for power bounded operators. Indiana Univ Math J 59:1121–1144
- Mustafayev HS (2014) The behavior of the orbits of power bounded operators. Oper Matrices 8:975–997
- Nagel R (1974) Ergodic and mixing properties of linear operators. Proc R Ir Acad 74:245–261
- Nagy B, Zemánek J (1999) A resolvent condition implying power boundedness. Stud Math 134:143–151
- Nair R (1993) On polynomials in primes and J. Bourgain's circle method approach to ergodic theorems, II. Stud Math 105:207–233
- Nasri-Roudsari D, Nessel RJ, Zeler R (1995) Resonance principles with applications to mean ergodic theorems and projection operators. Acta Math Hungar 68:269–285
- Nevanlinna O (1997) On the growth of the resolvent operators for power bounded operators. In: Linear operators, Banach Center Publications, vol 38. IMPAN, Warsaw, pp 247–264
- Nevanlinna O (2001) Resolvent conditions and powers of operators. Stud Math 145:113–134
- Ng ACS, Seifert D (2020) Optimal rates of decay in the Katznelson-Tzafriri theorem for operators on Hilbert spaces. J Funct Anal 279(12):108799, 21 p
- Oxtoby J (1952) Ergodic sets. Bull Am Math Soc 58:116–136
- Pazy A (1972) On the applicability of Lyapunov's theorem in Hilbert space. SIAM J Math Anal 3:291–294
- Petersen K (1983) Another proof of the existence of the ergodic Hilbert transform. Proc Am Math Soc 88:39–43
- Petersen K (1996) Ergodic theorems and the basis of science. Synthese 108:171–183
- Radjavi H, Tam P-K, Tan K-K (2003) Mean ergodicity for compact operators. Stud Math 158:207–217
- Reich S (1973) Iterative solution of linear operator equations in Banach spaces. Atti Accad Naz Lincei Rend Cl Sci Fis Mat Nat 54(8):551–554
- Riesz F (1938) Some mean ergodic theorems. J Lond Math Soc 13:274–278
- Ritt RK (1953) A condition that $\lim_{n \rightarrow \infty} n^{-1} T^n = 0$. Proc Am Math Soc 4:898–899
- Robinson EA (1960) Sums of stationary random variables. Proc Am Math Soc 11:7779
- Rokhlin VA (1948) A "general" measure-preserving transformation is not mixing (Russian). Doklady Akad Nauk SSSR (NS) 60:349–351
- Rokhlin VA (1961) Exact endomorphisms of Lebesgue spaces (Russian), Izv. Akad Nauk SSSR Ser Mat 25:499–530; English translation: Am Math Soc Transl 39(2), (1964), 1–36
- Rosenblatt JM (1994) Norm convergence in ergodic theory and the behavior of Fourier Transforms. Canadian J Math 46:184–199
- Rosenblatt JM, Wierdl M (1995) Pointwise ergodic theorems via harmonic analysis. In: Ergodic theory and its connections with harmonic analysis, London Mathematical Society, Lecture notes series, 205. Cambridge University Press, Cambridge, UK, pp 3–151
- Rozendaal J, Veraar M (2018) Sharp growth rates for semigroups using resolvent bounds. J Evol Equ 18:1721–1744
- Rudolph DJ (1994) A joinings proof of Bourgain's return time theorem. Ergodic Theory Dyn Syst 14:197–203
- Ryll-Nardzewski C (1975) Topics in ergodic theory. In Probability winter school, Karpacz 1975, Lecture notes in Mathematics, vol 472. Springer, Berlin, pp 131–156
- Satō R (1979) The Hahn-Banach theorem implies Sine's mean ergodic theorem. Proc Am Math Soc 77:426
- Satō R (1980) On the Blum-Hanson theorem in L_p . Math J Okayama Univ 22:27–32
- Seifert D (2016) Rates of decay in the classical Katznelson-Tzafriri theorem. J Anal Math 130:329–354
- Shields A (1978) On Möbius bounded operators. Acta Sci math (Szeged) 40:371–374

- Sine R (1970) A mean ergodic theorem. *Proc Am Math Soc* 24:438–439
- Sine R (1976) A note on the ergodic properties of homeomorphisms. *Proc Am Math Soc* 57:169–172
- Sine R (1991) Constricted systems. *Rocky Mountain J Math* 21:1373–1383
- Stampfli J (1972) A local spectral theory for operators III. *Trans Am Math Soc* 168:133–151
- Stein EM (1982) The development of square functions in the work of A. Zygmund. *Bull Am Math Soc (NS)* 7: 359–376
- Strikwerda J, Wade B (1997) A survey of the Kreiss matrix theorem for power bounded families of matrices and its extensions. In: *Linear operators*, Banach Center Publication, vol 38. IMPAN, Warsaw, pp 339–360
- Sucheston L (1976) Problems. In: *Probability in Banach spaces* (Oberwolfach 1975), Springer lecture notes in Mathematics, vol 526. Springer, Berlin/New York, pp 285–290
- Suciu L (2016) Estimations of the operator resolvent by higher order Cesàro means. *RM* 69:457–475
- Suciu L, Zemánek J (2013) Growth conditions and Cesàro means of higher order. *Acta Sci Math (Szeged)* 79: 545–581
- Sz-Nagy B, Foiaş C (1967) *Analyse harmonique des opérateurs de l'espace de Hilbert*, Masson, Paris; Akadémiai Kiadó, Budapest. English translation: *Harmonic analysis of operators on Hilbert space*, North-Holland, Amsterdam and Akadémiai Kiadó, Budapest, 1970. Second edition (with coauthors Bercovici and Kérchy), Universitext, Springer, New York, 2010
- Tempelman A (1974) Ergodic theorems for amplitude modulated random fields. *Litovsk Mat Sbor* 14: 221–229, (in Russian). Engl. transl.: *Lithuan. Math. Transl* 14 (1975), 698–704
- ter Elst AFM, Müller V (2017) A van der Corput-type lemma for power bounded operators. *Math Z* 285: 143–158
- Tomilov Y (2001) A resolvent approach to stability of operator semigroups. *J Operator Th* 46:63–98
- Tomilov Y, Zemánek J (2004) A new way of constructing examples in operator ergodic theory. *Math Proc Camb Philos Soc* 137:209–225
- Van Castren J (1980) A problem of Sz. Nagy. *Acta Sci math* (Szeged) 42:189–194
- van Neerven J (1996) The asymptotic behaviour of semigroups of linear operators, *Operator theory: advances and applications*, vol 88. Birkhäuser, Basel
- Visser C (1938) On the iteration of linear operations in a Hilbert space. *Neder Akad Wetensch* 41:487–495
- Volný D (2018) Martingale-coboundary representation for stationary random fields. *Stoch Dyn* 18(2):1850011. 18 pp
- von Neumann J (1932) Proof of the quasi-ergodic hypothesis. *Proc Natl Acad Sci U S A* 18:70–82
- Vũ QP (1992a) A short proof of the Y. Katznelson's and L. Tzafriri's theorem. *Proc Am Math Soc* 115: 1023–1024
- Vũ QP (1992b) Theorems of Katznelson-Tzafriri type for semigroups of operators. *J Funct Anal* 103:74–84
- Weiss G (1989) Weakly ℓ^p -stable linear operators are power stable. *Int J Systems Sci* 20:2323–2328
- Weyl H (1910) Über die Gibbs'sche Erscheinung und verwandte Konvergenzphänomene. *Rend Circ Mat Palermo* 30:377–407; reprinted In: *Weyl, Gesammelte Abhandlungen*, Band I, Springer-Verlag, Berlin/Heidelberg/New York, 1968, pp 321–353
- Weyl H (1916) Über die Gleichverteilung von Zahlen mod. Eins, *Math Ann* 77:313–352; reprinted In: *Weyl, Gesammelte Abhandlungen*, Band I, Springer-Verlag, Berlin/Heidelberg/New York, 1968, 563–599
- Wiener N, Wintner A (1941) Harmonic analysis and ergodic theory. *Am J Math* 63:415–426
- Wierdl M (1988) Pointwise ergodic theorem along the prime numbers. *Israel J Math* 64:315–336
- Wintner A (1945) The linear difference equation of first order for angular variables. *Duke Math J* 12:445–449
- Yoshimoto T (1993) On the speed of convergence in the (C, α) uniform ergodic theorem for quasi-compact operators. *J Math Anal Appl* 176:413–422
- Yoshimoto T (1996) On the convergence rate in the uniform ergodic theorem. *J Math Anal Appl* 200:149–161
- Yoshimoto T (1998) Uniform and strong ergodic theorems in Banach spaces. III *J Math* 42:525–543
- Yosida K (1938) Mean ergodic theorem in Banach spaces. *Proc Imp Acad Tokyo* 14:292–294
- Yosida K, Kakutani S (1938) Applications of mean ergodic theorems to the problems of Markoff process. *Proc Imp Acad Tokyo* 14:333–339
- Yosida K, Kakutani S (1941) Operator-theoretical treatment of Markoff process and mean ergodic theorem. *Ann Math* 42:188–228
- Zaharopol R (1986) Mean ergodicity of power-bounded operators in countably order complete Banach lattices. *Math Z* 192:81–88
- Zemánek J (1994) On the Gelfand-Hille theorems. In: *Functional analysis and operator theory*, Banach Center Publication 30, Polish Academy of Sciences, Institute Mathematics, Warsaw, pp 369–385
- Zsidó L (2007) Weak mixing properties of vector sequences. In: *The extended field of operator theory*, Operator theory advances and application, vol. 171, Birkhäuser, Basel, pp 361–388

Books

- Dunford N, Schwartz JT (1958) *Linear operators*, part I, Interscience. Wiley, New York
- Eisner T, Farkas B, Haase M, Nagel R (2015) *Operator theoretic aspects of ergodic theory*, Graduate texts in mathematics, vol 272. Springer, Cham
- Emel'yanov E (2007) *Non-spectral asymptotic analysis of one-parameter operator semigroups*, Operator theory: advances and applications, vol 173. Birkhäuser, Basel
- Engel KJ, Nagel R (2000) *One-parameter semigroups for linear evolution equations*, Graduate texts in mathematics, vol 194. Springer, New York

- Foguel SR (1969) The ergodic theory of Markov processes. Van-Nostrand, New York
- Goldstein J (1985) Semigroups of linear operators and applications, Oxford mathematical monographs. The Clarendon Press, Oxford University Press, New York. Second ed. Dover Publications, Mineola, NY, 2017.
- Halmos PR (1960) Lectures on ergodic theory. Chelsea Publishing Co., New York
- Hille E, Phillips RS (1957) Functional analysis and semigroups, rev edn., vol 31. American Mathematical Society Colloquium Publications, American Mathematical Society, Providence
- Krengel U (1985) Ergodic theorems. de Gruyter, Berlin
- Riesz F, Sz-Nagy B (1955) Leçons d'analyse fonctionnelle (French), 3rd edn. Gauthier-Villars, Paris; Akadémiai Kiadó, Budapest. English translation: Functional analysis, Dover Publications, Inc., New York, 1990
- Schaefer HH (1974) Banach lattices and positive operators, Die Grundlehren der mathematischen Wissenschaften, Band 215. Springer, New York/Heidelberg
- Yosida K (1980) Functional analysis, Sixth edition, Grundlehren der Mathematischen Wissenschaften, vol 123. Springer, Berlin/New York

Dynamical Systems and C^* -Algebras

T. Giordano and H.-C. Liao

University of Ottawa, Ottawa, ON, Canada

Article Outline

Glossary

Definition of the Subject

Introduction

Topological Orbit Equivalence

Mean Dimension, Small Boundary Property, and

Classification of C^* -Algebras

Boundary Actions and C^* -Simplicity

C^* -Simplicity and Unique Trace Property

References

Glossary

AF algebra An inductive limit of finite-dimensional C^* -algebras.

AF relation An inductive limit of finite equivalence relations.

Affable equivalence relation An étale equivalence relation that is orbit equivalent to an AF relation.

Almost finite actions An action that admits disjoint towers such that each tower has a shape that is almost invariant under translations and the union of the towers almost covers the entire space.

Bratteli diagram An infinite graph with a sequence of finite sets V_n of vertices and a sequence of finite sets E_n of edges connecting the vertices of V_{n-1} and V_n .

Bratteli–Vershik transformation A homeomorphism of the path space of a Bratteli diagram that can be viewed as a generalized odometer.

C^* -algebra Algebra of operators on a Hilbert space that is closed under adjoint and closed in the norm topology.

Conjugacy Two actions $G \curvearrowright X_1$ and $G \curvearrowright X_2$ are conjugate if there is a homeomorphism $F : X_1 \rightarrow X_2$ such that $F(gx) = gF(x)$ for every $x \in X_1$; and $g \in G$.

Continuous orbit equivalence Two group actions are continuously orbit equivalent if they are topologically orbit equivalent and the associated orbit cocycles are continuous.

Countable amenable group A countable group that admits finite subsets that are almost invariant under translation.

Dimension group An ordered abelian group that is an inductive limit of simplicial groups. It is a complete invariant for AF algebras.

Étale equivalence relation An equivalence relation on a topological space with a topology such that the map $(x, y) \mapsto y$ is a local homeomorphism. For an action $G \curvearrowright X$ the associated equivalence relation is given by $x \sim y \Leftrightarrow gx = y$ for some $g \in G$.

Free action An action $G \curvearrowright X$ whose stabilizer subgroup $\{g \in G : gx = x\}$ is trivial for every $x \in X$.

G -boundary A topological space X equipped with a minimal G -action such that the induced action of G on the space of regular Borel G -invariant probability measures on X is proximal.

Isomorphism Two actions $G_1 \curvearrowright X_1$ and $G_2 \curvearrowright X_2$ are isomorphic if there is a homeomorphism $F : X_1 \rightarrow X_2$ and a group isomorphism $\alpha : G_1 \rightarrow G_2$ such that $F(gx) = \alpha(g)F(x)$ for every $x \in X_1$.

Minimal action An action $G \curvearrowright X$ where every orbit $\{gx : g \in G\}$ is dense in X .

Proximal action An action $G \curvearrowright X$ where for every pair x_1, x_2 in X there is a net t_i in G such that $\lim_i t_i x_1 = \lim_i t_i x_2$.

Reduced group C^* -algebras A C^* -algebra constructed from the left regular representation of a group.

Shape The finite set $\{g_1, \dots, g_n\}$ associated to a tower $\bigcup_{i=1}^n g_i V$.

Simple C^* -algebra A C^* -algebra that has no nontrivial norm-closed ideals.

Strong orbit equivalence Two group actions are strongly orbit equivalent if they are topologically orbit equivalent and the associated orbit cocycles have at most one point of discontinuity.

Topological dynamical system A continuous group homomorphism from a topological group G into the homeomorphism group of a topological space X . It is also called a G -action on X , denoted by $G \curvearrowright X$, and X is called a G -space.

Topological orbit equivalence for actions Two group actions are topologically orbit equivalent if there is a homeomorphism F between the underlying topological spaces which maps bijectively orbits onto the orbits.

Topological orbit equivalence Two étale equivalence relations $\mathcal{R}_1$ and $\mathcal{R}_2$ are topologically orbit equivalent if there is a homeomorphism F between the underlying topological spaces such that $F \times F(\mathcal{R}_1) = \mathcal{R}_2$.

Topologically free action An action where the set of points x for which the stabilizer subgroup is trivial is dense in X .

Tower A finite disjoint union $\bigcup_{i=1}^n g_i V$ of translate of a set V in a G -space.

Unique trace property A C^* -algebra has the unique trace property if it has exactly one tracial state.

Definition of the Subject

In (1936), Murray and von Neumann showed that factors, the building blocks of von Neumann algebras, are of three different types. Defining the group measure space construction, which associates a von Neumann algebra to a measurable dynamical system, they constructed examples of factors of type II and III. Since then the interplay between von Neumann and C^* -algebras and the theory of measurable and of topological dynamical systems has continued to grow.

In this survey we review some recent developments between C^* -algebras and topological dynamics.

Introduction

The appearance of operators on a Hilbert space in the study of dynamical systems can be traced back to von Neumann's mean ergodic theorem (von Neumann 1932). It is perhaps not surprising that von Neumann also laid the foundation of operator algebras. In the seminal paper (Murray and von Neumann 1936), the authors introduced the *group measure space construction*, which associates a (von Neumann) algebra to a measure-preserving group action on a measure space. Since then the field of dynamical systems has never ceased to provide interesting and important examples in operator algebras. Meanwhile, the study of orbit equivalence in dynamics, initiated by Dye (1959), was largely inspired by Murray and von Neumann's classification of hyperfinite II_1 factors (Murray and von Neumann 1943). This connection extends to the much broader setting of amenable operator algebras and amenable equivalence relations, as seen in the Connes–Haagerup classification of amenable simple von Neumann algebras (Connes 1976; Haagerup 1987) and the Connes–Feldman–Weiss theorem for countable amenable Borel equivalence relations (Connes et al. 1982). Beyond the amenable case, the study of von Neumann algebras has led to remarkable results in orbit equivalence rigidity (see, e.g., Ioana (2013, 2018) for surveys on this topic).

Another striking analogy between the two fields is the interplay between measure and topology. Needless to say, measurable dynamics and topological dynamics are two branches in dynamical systems that are intimately related. On the operator algebras side, von Neumann algebras can be seen as noncommutative measure-theoretic objects, while C^* -algebras have a more topological flavor. From this viewpoint, it is only natural that the research on linkage between topological dynamics and C^* -algebras has seen a lot of progress in the past decades, and this is what the present survey focuses on.

In the section “[Topological Orbit Equivalence](#)” we review the theory of topological orbit equivalence for actions of finitely generated abelian groups on the Cantor set. Similar to the measurable case, the language of equivalence relations is

pivotal in classifying such actions. Moreover, the classification of AF (approximately finite) relations, which provides the base for classifying more general actions, relies on the classification of AF C^* -algebras. This again parallels the influence von Neumann algebras had on the study of orbit equivalence in the measurable case. In this section, we also describe some of the recent results on continuous orbit equivalence rigidity.

In the section “[Mean Dimension, Small Boundary Property, and Classification of \$C^*\$ Algebras](#),” we discuss recent developments in classification of amenable C^* -algebras and the notion of “regularity.” The main theme here is to connect regularity at the dynamical level to the C^* -algebraic level via the crossed product construction. It turns out that certain natural and purely dynamical questions, such as the existence of Kakutani–Rokhlin type decomposition for actions of more general groups, have a direct impact on the study of the associated C^* -algebras. Dynamical ideas such as mean dimension and comparing sets by invariant measures are also crucial in this type of connections. At the same time, the study of regularity of crossed product C^* -algebras also led to interesting dynamical results, such as a new characterization of the small boundary property in terms of certain kind of tower decomposition.

Finally, the section “[Boundary Actions and \$C^*\$ -Simplicity](#)” is devoted to some of the recent breakthroughs that apply boundary actions to the study of group C^* -algebras. More precisely we review the notion of Furstenberg boundary, discuss how it can be realized by an operator algebraic construction (the so-called Hamana boundary), and outline how this insight led to solutions to many long outstanding questions about group C^* -algebras.

Topological Orbit Equivalence

In 1959, H. Dye (1959) introduced the notion of orbit equivalence and proved that any two ergodic finite measure-preserving transformations on a Lebesgue space are orbit equivalent. In Dye (1963), he had also conjectured that an arbitrary ergodic action of a discrete amenable group is

orbit equivalent to a $\mathbb{Z}$ -action. This conjecture was proved by Ornstein and Weiss (1980). The most general case was proved by Connes, Feldman, and Weiss (1982) by establishing that an amenable nonsingular countable equivalence relation $\mathcal{R}$ can be generated by a single transformation, or equivalently, is hyperfinite, that is, $\mathcal{R}$ is, up to a null set, a countable increasing union of finite equivalence relations.

For the Borel case, Weiss proved that actions of $\mathbb{Z}^n$ are (orbit equivalent to) hyperfinite Borel equivalence relations, whose classification was obtained by Dougherty, Jackson, and Kechris (1994). It is not yet known if an arbitrary Borel action of a discrete amenable group is orbit equivalent to a $\mathbb{Z}$ -action. More generally, hyperfiniteness was proved for Borel actions of any finitely generated group with polynomial growth by Jackson–Kechris–Louveau (2002) and of any countable abelian group by Gao and Jackson (2015). In a recent preprint (Conley et al. 2020), this result was extended to Borel actions of polycyclic groups and for free actions of a large class of solvable groups including the Baumslag–Solitar group $BS(1,2)$ and the lamplighter group.

In this chapter, we present some of the developments in topological orbit equivalence, reviewing in particular the classification up to (topological) orbit equivalence of minimal actions of finitely generated abelian groups on the Cantor set. For connected spaces, using a result of Sierpinski (see Kuratowski (1968), Ch V, 47, III), any orbit equivalence is also an isomorphism. Therefore, we consider only spaces that are totally disconnected. The strategy in the topological case follows the one used both in the measurable and in the Borel case:

1. Provide an invariant of (topological) orbit equivalence for topological dynamical systems, which include actions of countable groups and étale equivalence relations. The definitions of étale equivalence relations and of the invariant, which is an ordered abelian group, in the subsections “[Étale Equivalence Relations](#)” and “[Invariants of Étale Equivalence Relations](#).”

2. Consider the rich and tractable class of AF (approximately finite) equivalence relations on the Cantor set. In the subsection “[AF-Equivalence Relations and Bratteli–Vershik Transformations](#),” we recall their definition and that they can always be realized as tail equivalence on a Bratteli diagram. In the subsection “[The Bratteli–Vershik Model](#),” we give the definition of the Bratteli–Vershik system introduced by R. Herman, I. Putnam, and C. Skau (1992) and state their remarkable theorem: any Cantor minimal system is conjugate to a Bratteli–Vershik system. There exist now several detailed presentations and surveys of this very important result (see Remark 4). In the subsection “[The Classification up to Isomorphisms of AF-Equivalence Relations: The Bratteli–Elliott–Krieger Theorem](#),” we recall the classification up to isomorphism of minimal AF relations on the Cantor set.

In “[Orbit Equivalence of AF-Equivalence Relations](#),” we classify, up to orbit equivalence, minimal AF-equivalence relations. A key technical result for this step is the absorption theorem for minimal AF-relations. This theorem, which we describe in “[The Absorption Theorem](#),” states that a “small” extension of a minimal AF relation is orbit equivalent to the original AF relation.

3. Prove that equivalence relations arising from minimal actions of finitely generated abelian groups are *affable*, that is, orbit equivalent to an AF-equivalence relation. We describe these results in “[Orbit Equivalence of Minimal Actions of a Finitely Generated Abelian Group](#)”.

In “[A Topological Krieger Theorem: The Notion of Strong Orbit Equivalence](#),” we describe links between Cantor minimal systems and the C^* -algebras associated to them. We introduce the notion of strong orbit equivalence and describe a topological Krieger theorem. We mention also Jewett–Krieger-type realization results proved by Ormes (1997).

In “[Continuous Orbit Equivalence](#)” and “[Continuous Orbit Equivalence and \$C^*\$ -algebra](#),” we

review the notion of continuous orbit equivalence which is in a one-to-one correspondence with a strong notion of isomorphism of the associated C^* -algebras. This domain of research is very active and we hope our description, even if it is very short, will show its importance. In “ [\$\mathbb{Z}^d\$ -odometers](#)” and “[Cohomology of \$\mathbb{Z}^d\$ -odometers](#),” we review the results on $\mathbb{Z}^d$ -odometers.

In the measurable case, full groups were introduced in 1963 by H. Dye in his study of orbit equivalence (1959, 1963). Several different full groups can be associated to a topological dynamical system (on the Cantor set). In particular, the *topological full groups*, introduced in Giordano et al. (1999) and Tomiyama (1996) for $\mathbb{Z}$ -actions, have been intensively studied and outstanding results obtained. We will not present them here, as several remarkable surveys have recently been written (de Cornulier 2014; Matui 2017; Katzlinger 2019).

Let Y be a compact metric space and G be an infinite countable group represented on Y as a group of self-homeomorphisms. Recall that the topological dynamical system (Y, G) is minimal if any of the following equivalent conditions are satisfied:

- (a) The empty set $\emptyset$ and Y are the only G -invariant closed subsets of Y .
- (b) For every $y \in Y$, its G -orbit $\text{Orb}_G\{y\} = \{gy \mid g \in G\}$ is dense in Y .
- (c) For every nonempty open set $U \subset Y$, there exists a finite subset

$$\{g_1, g_2, \dots, g_k\} \subset G \quad \text{with} \quad \bigcup_{j=1}^k g_j U = Y.$$

The following example is a key model of minimal homeomorphism of the Cantor set.

Example 1 Let $\underline{a} = (a_n)_{n \geq 1}$, with $a_n \in \mathbb{N}$, $a_n \geq 2$, and X be the compact abelian group $\prod_{n \geq 1} \{0, 1, \dots, a_n - 1\}$ (under the operation of addition mod a_n at the n -th coordinate, with carry over to the right). Then the $\underline{a}$ -adic adding machine, given by $\varphi(x) = x + (1, 0, 0, \dots)$, $x \in X$, is a minimal homeomorphism of the Cantor set.

Cantor minimal systems form a very rich class of topological dynamical systems. For example, recall that Ellis (1960) proved, for any countable (infinite), discrete group, the existence of a free, action on a compact metric space and therefore on the Cantor set by the following G -equivariant version of Alexander–Urysohn result (see Giordano and de la Harpe (1997) for a proof of this result).

Proposition 1 *Let α be a continuous action of a countable group G on a metrizable compact space Y .*

Then there exists a continuous action $\tilde{\alpha}$ of G on the Cantor set X and a factor map $\chi : X \rightarrow Y$ (i.e., an equivariant continuous surjective map).

If, moreover, Y is an infinite set and α is a minimal action, then $\tilde{\alpha}$ can be chosen minimal.

Étale Equivalence Relations

We first recall the definition and the first properties of étale equivalence relations (for more details see, e.g., Renault (1980); Paterson (1999); Putnam (2010)) and recall the notions of isomorphism and orbit equivalence of étale equivalence relations.

Let X be a compact, metric space and $\mathcal{R}$ be a countable equivalence relation on X , that is, $\mathcal{R} \subset X \times X$ is an equivalence relation so that for all $x \in X$, each equivalence class $\mathcal{R}[x] = \{y \in X \mid (x, y) \in \mathcal{R}\}$ is at most countable.

Recall that $\mathcal{R}$ has a natural (principal) groupoid structure, with *unit space* equal to the diagonal $\Delta = \{(x, x) \in \mathcal{R} \mid x \in X\}$. More specifically, if $(x, y), (y, z) \in \mathcal{R}$, then $(x, y)(y, z) = (x, z)$ is the product of this *composable pair* and the inverse of $(x, y) \in \mathcal{R}$ is the pair (y, x) . The unit space Δ of $\mathcal{R}$ is by definition $\{(x, y)(x, y)^{-1} \mid (x, y) \in \mathcal{R}\} = \{(x, x) \mid x \in X\}$. The range map $r : \mathcal{R} \rightarrow X$ and the source map $s : \mathcal{R} \rightarrow X$ are defined by $r(x, y) = y$ and $s(x, y) = x$, respectively, where $(x, y) \in \mathcal{R}$ and both maps are surjective.

Assume that $\mathcal{R}$ is given a Hausdorff locally compact, second countable (hence metrizable) topology τ , so that the product of composable pairs (with the topology inherited from the product topology on $\mathcal{R} \times \mathcal{R}$) is continuous. Also, the inverse map on $\mathcal{R}$ is a homeomorphism. With this

structure, $(\mathcal{R}, \tau)$ is a locally compact (principal) groupoid, cf. Renault (1980).

Definition 1 (Étale equivalence relation). *The locally compact principal groupoid $(\mathcal{R}, \tau)$ on the compact metric space X is an étale equivalence relation if $r : \mathcal{R} \rightarrow X$ is a local homeomorphism, that is, for all $(x, y) \in \mathcal{R}$, there exists an open neighborhood $U(x, y) \in \tau$ of (x, y) so that $r(U(x, y))$ is open in X and $r : U(x, y) \rightarrow r(U(x, y))$ is a homeomorphism. In particular, r is an open map.*

Example 2 Let G be a countable discrete group acting freely on a locally compact metric space X . Let

$$\mathcal{R}_G = \{(x, gx) \mid x \in X, g \in G \subset X \times X\}$$

that is, the $\mathcal{R}_G$ -equivalence classes are simply the G -orbits. Topologize $\mathcal{R}_G$ by transferring the product topology on $X \times G$ to $\mathcal{R}_G$ via the bijection $(x, g) \rightarrow (x, gx)$. As G is discrete, $\mathcal{R}_G$ is an étale equivalence relation. (If G does not act freely, we get a bijection between $\mathcal{R}_G$ and a closed subset of $X \times G \times X$ by the map $(x, gx) \rightarrow (x, g, gx)$ and we transfer the product topology on $X \times G \times X$ to $\mathcal{R}_G$.)

Remark 1

- 1) This definition of *étaleness* is equivalent to the various definitions of an étale (or r -discrete) locally compact groupoid that can be found in the literature (e.g., Renault (1980), Definition 2.6 and Proposition 2.8 and Paterson (1999), Definitions 2.2.1 and 2.2.3). The existence of an (essential) unique Haar system consisting of counting measures follows from our definition (see Paterson (1999), Proposition 2.2.5).
- 2) A countable equivalence relation on the Cantor set may be given nonequivalent étale topologies τ_i , $i = 1, 2$, contrasting with the situation in the countable (standard) Borel equivalence relation setting, where the Borel structure is uniquely determined by the inclusion of $\mathcal{R}$ in $X \times X$.
- 3) A countable equivalence relation $\mathcal{R}$ on a compact metric space X cannot always be endowed

with an étale topology. For example, the equivalence relation on $X = [0, 1]$ given by

$$\mathcal{R} = \Delta \cup \{(0, 1), (1, 0)\}$$

cannot be endowed with an étale topology τ . If τ were an étale topology on $\mathcal{R}$, then there would exist an open neighborhood U of $(0, 1)$ in $\mathcal{R}$ such that $r : U \rightarrow r(U)$ is a local homeomorphism. As the unit space Δ is closed, $\{(0, 1), (1, 0)\}$ is open in $\mathcal{R}$, and therefore $U \cap \{(0, 1), (1, 0)\}$ is also open and the restriction of r to $U \cap \{(0, 1), (1, 0)\}$ is a local homeomorphism. Thus, either $r|_{\{(0, 1)\}}$ or $r|_{\{(0, 1), (1, 0)\}}$ is open in $[0, 1]$. Contradiction!

Example 3 Let X be a compact metrizable space and $(\mathcal{R}, \tau)$ be a compact étale equivalence relation on X . Then (see Giordano et al. (2004), Proposition 3.2),

- 1) τ is the relative topology τ_{rel} from $\mathcal{R} \subset X \times X$.
- 2) There exists $M \geq 1$ such that, for all $x \in X$, $\#\mathcal{R}[x] \leq M$, where $\#\mathcal{R}[x]$ is the cardinality of $\mathcal{R}[x]$.

Therefore, if $(\mathcal{R}, \tau)$ is an étale equivalence relation and $\mathcal{R}$ contains an infinite equivalence class, then τ is not the relative topology from $\mathcal{R} \subset X \times X$.

The following proposition generalizes the Feldman–Moore characterization of countable (standard) Borel equivalence relations given in Feldman and Moore (1977), Theorem 1):

Proposition 2 (Giordano et al. (2004), Proposition 2.3). *Let $(\mathcal{R}, \tau)$ be an equivalence relation on a zero-dimensional set X . Then there exists a countable group G of homeomorphisms of X such that $\mathcal{R} = \mathcal{R}_G$.*

Note that it is not always possible to find a group G acting freely on X and such that $\mathcal{R} = \mathcal{R}_G$ as shown by Hjorth and Molberg (2006).

There are two natural notions of equivalence between étale equivalence relations.

Definition 2 (*Isomorphism and Orbit Equivalence*). *Let $(X_1, \mathcal{R}_1, \tau_1)$ and $(X_2, \mathcal{R}_2, \tau_2)$ be two étale equivalence relations on compact, metrizable spaces X_1 and X_2 . Then*

- 1) $(X_1, \mathcal{R}_1, \tau_1)$ and $(X_2, \mathcal{R}_2, \tau_2)$ are orbit equivalent if there exists a homeomorphism $h : X_1 \rightarrow X_2$, called an orbit equivalence, such that

$$(x, y) \in \mathcal{R}_1 \text{ if and only if } (h(x), h(y)) \in \mathcal{R}_2.$$
- 2) If moreover $h : X_1 \rightarrow X_2$ can be chosen such that $h \times h : (\mathcal{R}_1, \tau_1) \rightarrow (\mathcal{R}_2, \tau_2)$ is a homeomorphism, $(X_1, \mathcal{R}_1, \tau_1)$ and $(X_2, \mathcal{R}_2, \tau_2)$ are isomorphic.

Proposition 3 *Let (X_1, G_1) and (X_2, G_2) be two free actions of countable discrete groups G_1 and G_2 on compact metrizable spaces X_1 and X_2 . Then*

- 1) (X_1, G_1) is orbit equivalent to (X_2, G_2) if and only if $\mathcal{R}_{G_1}$ and $\mathcal{R}_{G_2}$ are orbit equivalent.
- 2) If (X_1, G_1) is isomorphic to (X_2, G_2) (i.e., there exist a homeomorphism $h : X_1 \rightarrow X_2$ and a group isomorphism $\alpha : G_1 \rightarrow G_2$ such that $h(gx) = \alpha(g)(hx)$, for all $g \in G_1$ and $x \in X_1$), then $\mathcal{R}_{G_1}$ and $\mathcal{R}_{G_2}$ are isomorphic.

Remark 2 The converse of the implication (2) does not hold as shown by H. Dahl (2008), where she constructs free minimal actions of non-isomorphic locally finite groups which give isomorphic equivalence relation.

By a theorem of Sierpinski (Kuratowski (1968), Theorem 6, Ch. V, §47, III), orbit equivalence implies isomorphism in the following case:

Theorem 1 *Let (X_1, G_1) and (X_2, G_2) be two free actions of countable discrete groups G_1 and G_2 on compact metrizable spaces X_1 and X_2 .*

If the two actions are orbit equivalent and if X_1 is connected, then they are isomorphic.

Let $\mathcal{R} \subset X \times X$ be an étale equivalence relation. Then as defined by J. Renault (1980), a Borel measure μ on X is $\mathcal{R}$ -invariant if $\mu(r(\mathcal{U})) =$

$\mu(s(\mathcal{U}))$, for every bisection $\mathcal{U} \subset \mathcal{R}$. Recall that $\mathcal{U}$ is a bisection of an étale equivalence relation $(\mathcal{R}, \tau)$ if it is open and the restriction of the range and the source map to $\mathcal{U}$ are homeomorphisms. The set of all bisections forms a basis for the topology τ (see, e.g., Renault (1980)). Let us denote by $M(X, \mathcal{R})$ the compact convex cone of $\mathcal{R}$ -invariant probability measures on X . Using that $\mathcal{R} = \mathcal{R}_G$ for some countable group G , the following proposition follows from a straightforward adaptation of the proof for Cantor minimal systems (see Giordano et al. (1995), p. 80). A direct and detailed proof for étale equivalence relations can be found in Putnam (2010, Theorem 2.8).

Proposition 4 *Let $(X_1, \mathcal{R}_1)$ and $(X_2, \mathcal{R}_2)$ be two orbit equivalent étale equivalence relations on compact, metrizable spaces X_1 and X_2 . Then there exists a homeomorphism $h : X_1 \rightarrow X_2$ which implements a bijection between $M(X_1, \mathcal{R}_1)$ and $M(X_2, \mathcal{R}_2)$.*

Invariants of Étale Equivalence Relations

To an étale equivalence relation $(X, \mathcal{R})$ on a totally disconnected, compact, Hausdorff space, we associate two preordered abelian groups $D(X, \mathcal{R})$ and $D_m(X, \mathcal{R})$ which are invariants of, respectively, isomorphism and orbit equivalence.

By a *preordered* group we mean a countable abelian group G with a subsemigroup G^+ containing 0, called the positive cone, such that

$$(i) \ G^+ + G^+ \subset G^+, \quad (ii) \ G = G^+ - G^+.$$

If moreover $G^+ \cap (-G^+) = \{0\}$, then G is an *ordered* group.

An *order unit* for (G, G^+) is an element $u \in G^+$ such that for all $g \in G$, there is $n \in \mathbb{N}$ with $nu - g \in G^+$.

Recall that $M(X, \mathcal{R})$ denotes the convex set of all $\mathcal{R}$ -invariant probability measures on X .

Definition 3 *Let $(X, \mathcal{R})$ be an étale equivalence relation on a totally disconnected compact metrizable space X , and $C(X, \mathbb{Z})$ be the countable abelian group of continuous functions from X to $\mathbb{Z}$. Then*

- (1) $B(X, \mathcal{R})$ is the subgroup of $C(X, \mathbb{Z})$ generated by all functions $\chi_{r(\mathcal{U})} - \chi_{s(\mathcal{U})}$, for a compact bisection $\mathcal{U} \subset \mathcal{R}$.
- (2) $B_m(X, \mathcal{R}) = \{f \in C(X, \mathbb{Z}) \mid \int f d\mu = 0, \forall \mu \in M(X, \mathcal{R})\}$.

Remark 3

- (a) If $M(X, \mathcal{R}) = \emptyset$, then $B_m(X, \mathcal{R}) = C(X, \mathbb{Z})$.
- (b) If $\mu \in M(X, \mathcal{R})$, then by definition $\mu(r(\mathcal{U})) = \mu(s(\mathcal{U}))$ for any compact bisection $\mathcal{U} \subset \mathcal{R}$. Hence, $B(X, \mathcal{R})$ is a subgroup of $B_m(X, \mathcal{R})$.

Definition 4 Let $(X, \mathcal{R})$ be as above. We define the two preordered abelian groups:

- (1) $D(X, \mathcal{R}) = C(X, \mathbb{Z})/B(X, \mathcal{R})$, where

$$D(X, \mathcal{R})^+ = \{[f] \mid \exists g \in C(X, \mathbb{Z}), g \geq 0 \text{ such that } f - g \in B(X, \mathcal{R})\}.$$

- (2) $D_m(X, \mathcal{R}) = C(X, \mathbb{Z})/B_m(X, \mathcal{R})$, where

$$D_m(X, \mathcal{R})^+ = \{[f]_m \mid \exists g \in C(X, \mathbb{Z}), g \geq 0 \text{ such that } \int (f - g) d\mu = 0, \forall \mu \in M(X, \mathcal{R})\}.$$

If $(X, \mathcal{R})$ is an étale equivalence relation on a totally disconnected compact metrizable space X , then by Remark 3 (b), there is a canonical positive homomorphism π_m from $D(X, \mathcal{R})$ onto $D_m(X, \mathcal{R})$.

By integration, any $\mu \in M(X, \mathcal{R})$ defines a state I_μ on $D(X, \mathcal{R})$, that is, a positive homomorphism from $D(X, \mathcal{R})$ to $\mathbb{R}$ such that $I_\mu([1]) = 1$, and as any state on $D(X, \mathcal{R})$ is of this form, we have

Proposition 5 *Let X be a totally disconnected, compact metrizable space and $\mathcal{R}$ be an étale equivalence relation on X .*

Then there is a bijective correspondence between the set $M(X, \mathcal{R})$ of $\mathcal{R}$ -invariant probability measures on X and the set $S(D(X, \mathcal{R}), [1_X])$ of states on $(D(X, \mathcal{R}), D(X, \mathcal{R})^+, [1_X])$.

By Remark 3 (a), if $M(X, \mathcal{R}) = \emptyset$, then $D_m(X, \mathcal{R}) = 0$, and as the infinitesimal subgroup $\text{Inf}(D$

$(X, \mathcal{R})$) of $D(X, \mathcal{R})$ is $\{[f] \in D(X, \mathcal{R}) \mid \tau([f]) = 0 \text{ for all } \tau \in S(D(X, \mathcal{R}), D(X, \mathcal{R})^+, [\mathbf{1}_X])\}$, we get

Proposition 6 *Let $(X, \mathcal{R})$ be as above. Then*

- (1) *The infinitesimal subgroup of $D(X, \mathcal{R})$ is equal to $B_m(X, \mathcal{R})/B(X, \mathcal{R})$*
- (2) *$D(X, \mathcal{R})/\text{Inf}(D(X, \mathcal{R}))$ is canonically order isomorphic to $D_m(X, \mathcal{R})$*

Then we get (see, e.g., Giordano et al. (2018), Proposition 20.4.36 and 20.4.40)

Proposition 7 *If $h : X \rightarrow X'$ is an orbit map between two étale equivalence relations $(X, \mathcal{R})$ and $(X', \mathcal{R}')$ on totally disconnected compact metrizable spaces X and X' , then it induces a unital order isomorphism from $(D_m(X, \mathcal{R}), [\mathbf{1}_X]_m)$ onto $(D_m(X', \mathcal{R}'), [\mathbf{1}_{X'}]_m)$.*

Moreover if h implements an isomorphism between $(X, \mathcal{R})$ and $(X', \mathcal{R}')$, then it induces an order isomorphism from $D(X, \mathcal{R})$ onto $D(X', \mathcal{R}')$ preserving the order units.

Example 4

1. Let φ be a minimal homeomorphism of the Cantor space X and $\mathcal{R}_\varphi = \{(x, \varphi^n(x)) \mid x \in X, n \in \mathbb{Z}\}$ be the corresponding étale equivalence relation. Then one checks that $B_\varphi = \{f - f \circ \varphi \mid f \in C(X, \mathbb{Z})\}$ and therefore

$$D(X, \mathcal{R}_\varphi) = C(X, \mathbb{Z}) / \{f - f \circ \varphi \mid f \in C(X, \mathbb{Z})\}.$$

Putnam (1989) in his study of the C^* -algebra associated to (X, φ) showed that $D(X, \mathcal{R}_\varphi)$ is a dimension group (see below for the precise definition), and Poon (1989) showed that $D(X, \mathcal{R}_\varphi)$ is an ordered group if φ is topologically transitive. In a recent prepublication, D. Handelman and M. Boyle show that if φ is a homeomorphism of a compact metrizable zero-dimensional space X , then φ is chain recurrent if and only if $D(X, \mathcal{R}_\varphi)$ is an unperforated abelian group.

2. For $N \geq 2$, there are examples of minimal actions φ of $\mathbb{Z}^N$ on the Cantor set, such that $D(X, \mathcal{R}_\varphi)$

has torsion (see Gähler et al. (2013) or Matui (2008b)). Hence, it is not a torsion free group.

AF-Equivalence Relations and Bratteli–Vershik Transformations

The AF-equivalence relations (Renault 2003; Giordano et al. 2004) form a rich, but also tractable class of étale equivalence relations. The terminology AF comes from C^* -algebra theory and means approximately finite.

Definition 5 *An étale equivalence relation $\mathcal{R}$ on a space X is an AF-equivalence relation if X is a totally disconnected compact metrizable space and if $\mathcal{R}$ is the union of an increasing sequence of compact open subequivalence relations.*

For a totally disconnected, compact, Hausdorff space, the above definition of an AF-equivalence relation is equivalent to the definition of an approximately proper (AP) equivalence relation introduced by Renault (2003).

Let G be a countable group acting minimally and freely on the Cantor set X . If G is locally finite, then it is the union of an increasing sequence of finite groups $(G_n)_{n \geq 1}$. The associated étale equivalence relation $\mathcal{R}_G$ can be identified with $X \times G$ and is equal to the inductive $\varinjlim \mathcal{R}_{G_n} = \varinjlim (X \times G_n)$. As for $n \geq 1$, $\mathcal{R}_{G_n}$ is clearly a compact étale equivalence relation and $\mathcal{R}_{G_n} \subset \mathcal{R}_{G_{n+1}}$, it follows that $\mathcal{R}_G$ is an AF-equivalence relation. The converse is also true.

Theorem 2 (Giordano et al. (2004), Theorem 3.8). *Let G be a countable group acting minimally and freely on the Cantor set X . Then $\mathcal{R}_G$ is AF if and only if G is locally finite.*

AF-equivalence relations satisfy the following stability properties:

Proposition 8 (Giordano et al. (2004), Proposition 3.12).

- 1) *An inductive limit of a sequence of AF-equivalence relations on X is AF.*

- 2) Let $(\mathcal{R}, \tau)$ be an AF-equivalence relation and $\mathcal{R}' \subset \mathcal{R}$ be an open subequivalence relation. Then $(\mathcal{R}', \tau|_{\mathcal{R}'})$ is AF.

Let us now describe the fundamental example of an AF-equivalence relation. We begin with a Bratteli diagram (see Herman et al. 1992; Effros 1981). It is a locally finite, infinite directed graph which consists of a vertex set $V = \bigsqcup_{n=0}^{\infty} V_n$ and an edge set $E = \bigsqcup_{n=1}^{\infty} E_n$, where the V_n 's and the E_n 's are finite disjoint sets.

The edges in E_n connect vertices in V_{n-1} with vertices in V_n . If e connects $v \in V_{n-1}$ with $u \in V_n$, we write $s(e) = v$ and $r(e) = u$, where $s : E_n \rightarrow V_{n-1}$ and $r : E_n \rightarrow V_n$ are the source and range maps, respectively. We will assume that $s^{-1}(v) \neq \emptyset$ for all $v \in V$ and that $r^{-1}(v) \neq \emptyset$ for all $v \in V \setminus V_0$.

Given a Bratteli diagram (V, E) and a sequence $0 = m_0 < m_1 < m_2 < \dots$ in $\mathbb{Z}^+$, the *telescoping* of (V, E) to $\{m_n\}$ is the Bratteli diagram (V', E') , where $V'_n = V_{m_n}$ and $E'_n = E_{m_{n-1}+1} \circ \dots \circ E_{m_n}$, and the source and the range maps are as above.

To a Bratteli diagram (V, E) , we associate its infinite path space $X_{(V,E)}$ given by

$$\{(e_1, e_2, \dots) | e_i \in E_i, r(e_i) = s(e_{i+1}) \forall i \geq 1\}.$$

Clearly $X_{(V,E)} \subseteq \prod_{n=1}^{\infty} E_n$, and endowed with the relative topology, $X_{(V,E)}$ is a compact, metrizable, totally disconnected space.

To a finite path $p = (e_1, e_2, \dots, e_n)$ starting at $v_0 \in V_0$ is associated the *cylinder set* $U(p) = \{(f_1, f_2, \dots) \in X_{(V,E)} | f_i = e_i, i = 1, 2, \dots, n\}$. The cylinder sets are clopen sets and form a basis for the topology of $X_{(V,E)}$. If (V, E) is simple (see below for the definition), then $X_{(V,E)}$ has no isolated points, and so $X_{(V,E)}$ is a Cantor set. (We will in the sequel disregard the trivial case where the path space $X(V, E)$ is a finite set.)

For each $N \geq 0$, let

$$\begin{aligned} \mathcal{R}_N &= \{(e, f) \in X_{(V,E)} \times X_{(V,E)} | e_n \\ &= f_n \text{ for all } n > N\}. \end{aligned}$$

Endowed with the relative topology of $X_{(V,E)} \times X_{(V,E)}$, then $\mathcal{R}_N$ is a compact étale equivalence relation (hence each equivalence class is finite), and $\mathcal{R}_N$ is an open subset of $\mathcal{R}_{N+1}$, for all $N \geq 0$.

Let $\mathcal{R} = \bigcup_{N \geq 0} \mathcal{R}_N$, and give $\mathcal{R}$ the inductive limit topology. This means that a sequence $\{(x_n, y_n)\}$ in $\mathcal{R}$ converges to (x, y) in $\mathcal{R}$ if and only if $\{x_n\}$ converges to x , $\{y_n\}$ converges to y (in X), and, for some N , (x_n, y_n) is in $\mathcal{R}_N$ for all but finitely many n . Then $\mathcal{R}$ is an étale equivalence relation which we denote by $AF(V, E)$.

If (V', E') is a Bratteli diagram obtained from (V, E) by telescoping, then the étale equivalence relations $AF(V, E)$ and $AF(V', E')$ are naturally isomorphic.

The Bratteli diagram (V, E) is *simple* if there exists a telescoping of (V, E) such that the resulting Bratteli diagram (V', E') has full connection between all consecutive levels. Then (V, E) is simple if and only if every $AF(V, E)$ -equivalence class is dense in $X_{(V,E)}$.

The above example is in fact the general case. Indeed, we have

Theorem 3 (Giordano et al. (2004), Theorem 3.9). *Let $\mathcal{R}$ be an AF-equivalence relation on a totally disconnected compact metrizable space X .*

Then, there exists a Bratteli diagram (V, E) such that $\mathcal{R}$ is isomorphic to the AF-equivalence relation $AF(V, E)$ on $X_{(V,E)}$.

Moreover, $(X, \mathcal{R})$ is minimal if and only if the Bratteli diagram (V, E) is simple.

Recall that the dimension group $G(V, E)$ associated to a Bratteli diagram (V, E) is the inductive limit of the sequence of the simplicial groups $(\mathbb{Z}^{|V_n|}, (\mathbb{Z}^+)^{|V_n|})$. Dimension groups were introduced by G. A. Elliott (1976) in his classification of AF C^* -algebras to denote any ordered group isomorphic to an inductive limit of a sequence of simplicial groups. The following result of Effros, Handelman, and Shen (1980) provides an abstract characterization of dimension groups.

Theorem 4 *A countable, ordered abelian group (G, G^+) is a dimension group if and only if*

- 1) it is unperforated, i.e. for every $g \in G$ and every positive integer n , if $ng \in G^+$, then $g \in G^+$, and
- 2) It has the Riesz interpolation property, that is, for any $a_1, a_2, b_1, b_2 \in G$ with $a_1, a_2 \leq b_1, b_2$, there exists $c \in G$ with $a_i \leq c \leq b_j$, for $i, j = 1, 2$.

Note that a dimension group being unperforated is torsion free. Recall that an *order ideal* in an ordered group (G, G^+) is a subgroup $\mathcal{I}$ such that $\mathcal{I} = \mathcal{I}^+ - \mathcal{I}^+$, where $\mathcal{I}^+ = \mathcal{I} \cap G^+$, and if $0 \leq a \leq b$, with $a \in G$ and $b \in \mathcal{I}$, then $a \in \mathcal{I}$, and that an ordered group G is *simple* if its only order ideals are $\{0\}$ and G .

We then have (see Herman et al. 1992):

Theorem 5 For a Bratteli diagram (V, E) and the associated AF-equivalence relation $AF(V, E)$ on $X_{(V, E)}$, the group $D(X_{(V, E)}, AF(V, E))$ is the dimension group $G(V, E)$ associated to (V, E) . It is simple if and only if $AF(V, E)$ is minimal.

The Bratteli–Vershik Model

Let (V, E) be a Bratteli diagram with $V_0 = \{v_0\}$. In (Herman et al. 1992), Herman, Putnam and Skau introduce the notion of an *ordered Bratteli diagram* by defining a partial order $\geq$ on E so that edges $e, e' \in E$ are comparable if and only if $r(e) = r(e')$. They then show that any simple Bratteli diagram can be *properly ordered*, that is, there exists a partial order on E with a unique min path $x_{\min} = (e_1, e_2, \dots)$ and a unique max path $x_{\max} = (f_1, f_2, \dots)$ in $X_{(V, E)}$ (i.e., e_i is a minimal edge and f_i a maximal one for all $i = 1, 2, \dots$).

As a properly ordered Bratteli diagram is simple, its path space $X_{(V, E)}$ is a Cantor set. Then in (Herman et al. 1992), they defined a generalized odometer on $X_{(V, E)}$ as follows:

Definition 6 Let $X_{(V, E)}$ be the path space of a properly ordered Bratteli diagram. Then the Bratteli–Vershik transformation $\varphi_{(V, E)}$ on $X_{(V, E)}$ is defined by: if $x = (e_1, e_2, \dots) \in X_{(V, E)}$ and $x \neq x_{\max}$, then $\varphi_{(V, E)}(x)$ is the successor of x in the lexicographic ordering and $\varphi_{(V, E)}(x_{\max}) = x_{\min}$.

Then they proved the following remarkable result:

Theorem 6 (Herman et al. 1992). Let (X, φ) be a Cantor minimal system. Then there exists a properly ordered Bratteli diagram $(V, E, \geq)$ such that the associated Bratteli–Vershik system $(X_{(V, E)}, \varphi_{(V, E)})$ is conjugate to (X, φ) .

Remark 4

- a) Vershik (1981, 1982) used sequences of refining measurable partitions of a measurable space to realize ergodic automorphisms of a measured space.
- b) There are several detailed presentations and surveys of the results of Herman et al. (1992), see, for example, Skau (2000), Durand (2010); Putnam (2010); Bezuglyi and Karpel (2016); Putnam (2018) and the monograph (Durand and Perrin 2022).
- c) The case of aperiodic Cantor systems was studied in Bezuglyi et al. (2005) and Medynets (2006).

Let $(X_{(V, E)}, \varphi_{(V, E)})$ be the Bratteli–Vershik system associated to a properly ordered Bratteli diagram $(V, E, \geq)$, and let $AF(V, E)$ be the associated AF-equivalence relation on $X_{(V, E)}$. By construction $AF(V, E) \subset \mathcal{R}_{\varphi_{(V, E)}}$ and by the techniques developed in Putnam (2010), we have

Theorem 7 (Herman et al. (1992), Cor. 5.4). Let $(X_{(V, E)}, \varphi_{(V, E)})$ be the Bratteli–Vershik system associated to a properly ordered Bratteli diagram $(V, E, \geq)$, and let $AF(V, E)$ be the associated AF-equivalence relation on $X_{(V, E)}$.

Then $D(X, \varphi)$ is order isomorphic to $D(X_{(V, E)}, AF(V, E))$ with distinguished order units.

Using now the Effros–Handelman–Shen Theorem (Effros et al. 1980), we obtain

Theorem 8 (Herman et al. (1992), Cor. 6.3). Let (X, φ) be a Cantor minimal system. Then $(D(X, \varphi), D(X, \varphi)^+, [1_X])$ is a unital, simple, acyclic (i.e., not order isomorphic to $\mathbb{Z}$) dimension group. Furthermore, if (G, G^+) is any simple, acyclic dimension group with distinguished order unit

u , there exists a Cantor minimal system (X, φ) so that $(G, G^+, u) \cong (D(X, \varphi), D(X, \varphi)^+, [\mathbf{1}_X])$.

The Classification up to Isomorphisms of AF-Equivalence Relations: The Bratteli–Elliott–Krieger Theorem

For AF-equivalence relations, the invariants introduced in Definition 4 have not only a well-known structure, but they also form complete sets of invariants of AF-equivalence relations on the Cantor set up to isomorphism and in the minimal case up to orbit equivalence. This result is based on the famous Bratteli–Elliott classification of AF C^* -algebras (Bratteli 1972; Elliott 1976) and on Krieger’s dynamical version (Krieger 1979).

Theorem 9 *Let $(X_1, \mathcal{R}_1)$ and $(X_2, \mathcal{R}_2)$ be two AF-equivalence relations on the Cantor set. Then they are isomorphic if and only if there exists an unital order group isomorphism from $D(X_1, \mathcal{R}_1)$ onto $D(X_2, \mathcal{R}_2)$.*

The Absorption Theorem

The absorption theorem is the key tool used to classify up to orbit equivalence minimal AF-equivalence relations and étale equivalence relations associated to minimal actions on the Cantor set of finitely generated abelian groups. This result shows that a “small” extension of a minimal AF-equivalence relation is orbit equivalent to the original one.

There are three different versions of the absorption theorem. The first version proved in Giordano et al. (2004) had to be generalized in Giordano et al. (2008) for the classification up to orbit equivalence of free minimal actions of $\mathbb{Z}^2$. It had to be strengthened again in order to generalize the results for minimal $\mathbb{Z}^2$ -actions to minimal $\mathbb{Z}^d$ -actions. This final version below is due to H. Matui (2008a).

To state the theorem, let us introduce the terminology we will need to define the notion of a “small extension.” If $(X, \mathcal{R})$ is an étale equivalence relation on the Cantor set X , then a closed subset Y of X is

1. $\mathcal{R}$ -étale if the restriction $\mathcal{R} \cap (Y \times Y)$ of $\mathcal{R}$ to Y , denoted by $\mathcal{R}|_Y$, is an étale equivalence relation in the relative topology;

2. $\mathcal{R}$ -thin if $\mu(Y) = 0$ for all $\mathcal{R}$ -invariant probability measures μ on X .

Remark 5

- 1) In Lemma 2.5 of Phillips (2005), the author gives a measure-free definition of an $\mathcal{R}$ -thin set.
- 2) If $B = (V, E)$ is a Bratteli diagram and $C = (W, F)$ a subdiagram (i.e., $W \subset V$, $F \subset E$, and $W_0 = V_0 = \{v_0\}$), then the closed subset X_C of X_B is an étale subset, and Theorem 3.11 of Giordano et al. (2004) shows that any $\mathcal{R}$ -étale subset of an AF-equivalence relation $\mathcal{R}$ is of this form.

Theorem 10 (Matui (2008a), Theorem 3.2). *Let $(X, \mathcal{R})$ be a minimal AF-equivalence relation on the Cantor set X and let $Y \subset X$ be a closed, $\mathcal{R}$ -étale and $\mathcal{R}$ -thin subset. Suppose that an AF-equivalence relation $\mathcal{Q} \subset Y \times Y$ is an étale extension of $\mathcal{R}|_Y$ (i.e., $\mathcal{R}|_Y \subset \mathcal{Q}$ and the inclusion map is continuous).*

If $\tilde{\mathcal{R}} = \mathcal{R} \vee \mathcal{Q}$ denotes the equivalence relation on X generated by $\mathcal{R}$ and $\mathcal{Q}$, then there exists a homeomorphism $h : X \rightarrow X$ such that

- (i) $(h \times h)(\tilde{\mathcal{R}}) = \mathcal{R}$. In particular, $\tilde{\mathcal{R}}$ is orbit equivalent to $\mathcal{R}$.
- (ii) $h(Y)$ is a closed, $\mathcal{R}$ -étale, and $\mathcal{R}$ -thin subset.
- (iii) $h|_Y \times h|_Y : Y \times Y \rightarrow h(Y) \times h(Y)$ is a homeomorphism from $\mathcal{Q}$ to $\mathcal{R}|_{h(Y)}$.

In the three versions of the absorption theorem, the strategy of the proof consists of constructing countable disjoint replicas of $\mathcal{R}|_Y$ and $\mathcal{Q}$, and then using the following strong version of Bratteli–Elliott–Krieger’s result.

Lemma 1 (Giordano et al. (2004), Lemma 4.15). *For $i = 1, 2$, let $(X_i, \mathcal{R}_i)$ be two isomorphic minimal AF-equivalence relations on the Cantor sets, and $Z_i \subset X_i$ be $\mathcal{R}_i$ -thin, and $\mathcal{R}_i$ -étale closed subsets.*

Assume that there exists a homeomorphism $\alpha : Z_1 \rightarrow Z_2$, which implements an isomorphism between $\mathcal{R}_1|_{Z_1}$ and $\mathcal{R}_2|_{Z_2}$.

Then there exists an extension $\tilde{\alpha} : X_1 \rightarrow X_2$ of α which implements an isomorphism between $\mathcal{R}_1$ and $\mathcal{R}_2$.

Definition 7 (Giordano et al. (2004), Definition 4.1). Let $\mathcal{R}$ be a countable equivalence relation on the Cantor set X . We say that $\mathcal{R}$ is affable if it may be given a topology τ so that $(\mathcal{R}, \tau)$ is an AF-equivalence relation or equivalently if $\mathcal{R}$ is orbit equivalent to an AF-equivalence relation.

Remark 6 Keeping the notation of Theorem 10, and calling the extension $\tilde{\mathcal{R}}$ of the minimal AF-equivalence relation $\mathcal{R}$ a “small” extension, then the absorption theorem states in particular that a “small” extension of a minimal AF-equivalence is not only affable, but orbit equivalent to $\mathcal{R}$.

Example 5 Let us use the absorption theorem to prove, for any Cantor minimal system (X, φ) , the affability of its associated étale equivalence relation

$$\mathcal{R}_\varphi = \{(x, \varphi^k(x)), x \in X \text{ and } k \in \mathbb{Z}\}.$$

Let $(U_n)_{n \geq 1}$ be a decreasing sequence of clopen subsets of X , whose intersection is a single point y , and for $n \geq 1$, let $\mathcal{R}_n$ denote the equivalence relation on X generated by $\{(x, \varphi(x)) \mid x \in X \setminus U_n\}$. As φ is minimal and as the first return map of φ on U_n is continuous, we have that $\mathcal{R}_n$ is compact and open. As $(U_n)_{n \geq 1}$ forms a decreasing sequence of clopen sets, the sequence of the equivalence relations $(\mathcal{R}_n)_{n \geq 1}$ is increasing and their union $\mathcal{R}_y$ is an AF-relation. Every $\mathcal{R}_y$ -class is also a φ -orbit, except for the orbit of the point y and $\mathcal{R}_\varphi = \mathcal{R}_y \vee \{(y, \varphi(y))\}$.

Then $Y = \{y, \varphi(y)\}$ is evidently a closed and $\mathcal{R}_y$ -thin subset of X . As $(y, \varphi(y)) \notin \mathcal{R}_y$, we see that

$$\mathcal{R}_y|_Y = \Delta_Y = \{(y, y), (\varphi(y), \varphi(y))\} \subset Y \times Y$$

and Y is an $\mathcal{R}_y$ -étale subset of X . Let $Q = Y \times Y$. Clearly Q is an AF-equivalence relation on Y containing $\mathcal{R}_y|_Y$.

As $\mathcal{R}_y \vee Q = \mathcal{R}_\varphi$, by the absorption theorem $\mathcal{R}_y$ and $\mathcal{R}_\varphi$ are orbit equivalent, and therefore we have

Theorem 11 Any minimal $\mathbb{Z}$ -action on the Cantor set is affable.

Orbit Equivalence of AF-Equivalence Relations

The classification up to orbit equivalence of minimal AF-equivalence relation will be a corollary of the following theorem proved in 2010 by I. Putnam (2010).

Theorem 12 Let $(X, \mathcal{R})$ be a minimal AF-equivalence relation on the Cantor set X .

Then there exists a minimal AF-equivalence relation $\tilde{\mathcal{R}}$ on X , containing $\mathcal{R}$ and such that

- (1) $\tilde{\mathcal{R}}$ is orbit equivalent to $\mathcal{R}$.
- (2) the two unital dimension groups $D(X, \tilde{\mathcal{R}})$ and $D_m(X, \mathcal{R})$ are order isomorphic.

Sketch of the main steps of the proof of the theorem:

- (1) By the Effros–Handelman–Shen Theorem 4, there exists a Bratteli diagram (W, F) such that $G(W, F) = D(X, \mathcal{R}) / \text{Inf}(D(X, \mathcal{R})) = D_m(X, \mathcal{R})$.
- (2) The key technical part (and a real “tour de force”) of Ian Putnam’s proof is the construction of a “nice” Bratteli diagram (V, E) and of a graph homomorphism: $h : (V, E) \rightarrow (W, F)$ with the following properties:
 - (a) $G(V, E) = D(X_E, \mathcal{R}_E) = D(X, \mathcal{R})$.
 - (b) h induces a homeomorphism $h : X_E \rightarrow X_F$ such that $h \times h : \mathcal{R}_E \rightarrow \mathcal{R}_F$ is continuous and open.
 - (c) There exists a subdiagram Δ of the Bratteli diagram (V, E) such that $Y = X_\Delta$ is a closed, $\mathcal{R}_E$ -étale, $\mathcal{R}_E$ -thin subset of X_E and an AF-equivalence relation $Q \supset \mathcal{R}|_Y$ on $Y \times Y$ such that $\tilde{\mathcal{R}} = \mathcal{R} \vee Q$.

By the absorption theorem (Theorem 10), $\tilde{\mathcal{R}}$ is orbit equivalent to $\mathcal{R}$ and $h \times h(\tilde{\mathcal{R}}) = \mathcal{R}_F$. By transferring the topology of $\mathcal{R}_F$ via

$(h \times h)^{-1}, \tilde{\mathcal{R}}$ becomes an AF-equivalence relation containing $\mathcal{R}$ and whose unital dimension group is order isomorphic to $D_m(X_E, \mathcal{R})$.

As a consequence of Theorem 12 and Theorem 9, we obtain the classification up to orbit equivalence of AF-equivalence relations.

Corollary 1 (Putnam (2010), Cor. 3.2). *For a minimal AF-equivalence relation $(X, \mathcal{R})$ on the Cantor set X , its unital dimension group*

$$(D_m(X, \mathcal{R}), D_m(X, \mathcal{R})^+, [1_X])$$

is a complete invariant of orbit equivalence.

Remark 7

- 1) The classification of AF-equivalence relations was originally given in Giordano et al. (1995), as a consequence of the classification of minimal Cantor systems. The proof in Giordano et al. (1995) relies on nontrivial results in homological algebra. The new proof given by Ian Putnam is certainly the “right” one, as it avoids completely the use of homological algebra and is very dynamical in nature.
- 2) In Putnam (2010), the following generalized version of Theorem 12 is in fact proved:
Let $(X, \mathcal{R})$ be a minimal AF-equivalence relation and H be a subgroup of $\text{Inf}(D(X, \mathcal{R}))$ such that $D(X, \mathcal{R})/H$ is torsion free.
Then there exists a minimal AF-equivalence relation $\tilde{\mathcal{R}}$ on X , containing $\mathcal{R}$ and orbit equivalent to $\mathcal{R}$ and such that

$$D(X, \tilde{\mathcal{R}}) \cong D(X, \mathcal{R})/H.$$

Orbit Equivalence of Minimal Actions of a Finitely Generated Abelian Group

The complete classification of minimal AF-equivalence relations on the Cantor set up to orbit equivalence was given in Corollary 1. In this section, we extend this complete classification to minimal actions of finitely generated abelian groups on the Cantor set by showing that these

actions are *affable*, that is, orbit equivalent to an AF-equivalence relation. The key tool for this step is the absorption theorem for minimal AF relations (Theorem 10).

We first describe the case of minimal $\mathbb{Z}$ -actions studied in Giordano et al. (1995). Notice however that the proof of the classification up to orbit equivalence of minimal $\mathbb{Z}$ -actions given in Giordano et al. (1995) did not use the absorption theorem (see Remark 7).

We then discuss the affability of minimal actions of $\mathbb{Z}^d$ on the Cantor set. The case $d = 2$ was proved in Giordano et al. (2008) and the general case in Giordano et al. (2010). This proof of the affability is extended to minimal actions of finitely generated abelian groups by using a result that Ø. Johansen proved in his thesis (Johansen 1998).

By Theorem 14, any minimal $\mathbb{Z}$ -action on the Cantor set is affable. Therefore, we get

Theorem 13 (Giordano et al. (1995), Theorem 2.2). *For $i = 1, 2$, let $(X_i, \mathcal{R}_i)$ be the two minimal equivalence relations on the Cantor set, which are either AF or induced by a Cantor minimal system.*

Then the following statements are equivalent:

- (1) $(X_1, \mathcal{R}_1)$ and $(X_2, \mathcal{R}_2)$ are orbit equivalent.
- (2) The dimension groups $D_m(X_1, \mathcal{R}_1)$ and $D_m(X_2, \mathcal{R}_2)$ are order isomorphic by a map preserving the distinguished order units.
- (3) There exists a homeomorphism $F : X_1 \rightarrow X_2$ carrying $M(X_1, \mathcal{R}_1)$ onto $M(X_2, \mathcal{R}_2)$.

The equivalence between (1) and (2) is given by Corollary 1, and the implication from (3) to (2) follows directly from the definition of $B_m(X_i, \mathcal{R}_i)$ (Definition 3). By Proposition 4, (1) implies (3).

By Proposition 1.16 and Corollary 4.2 of Effros (1981), we then get

Corollary 2 (Giordano et al. (1995), Cor. 2.1). *Let (X_1, φ_1) and (X_2, φ_2) be two uniquely ergodic Cantor minimal systems. Then they are orbit equivalent if and only if $\{\mu_1(E) | E \subset X_1, \text{ clopen}\} = \{\mu_2(E) | E \subset X_2, \text{ clopen}\}$, where μ_i is the unique φ_i -invariant measure on X_i for $i = 1, 2$.*

Recall (see Giordano et al. (1995), p. 63 and Putnam et al. (1986) for a survey) that a Denjoy homeomorphism is an aperiodic homeomorphism of $\mathbb{S}^1$, which is not conjugate to a rigid rotation. By a *Denjoy system*, we mean a Denjoy homeomorphism restricted to its unique invariant Cantor set. A Denjoy system is uniquely ergodic.

Let μ be the invariant probability measure of a uniquely ergodic Cantor minimal system (X, φ) , and G be the subgroup $\{\mu(E); E \subset X, \text{ clopen}\} + \mathbb{Z}$ of $\mathbb{R}$. By Corollary 2,

$$(D_m(X, \mathcal{R}_\varphi), [\mathbf{1}_X]) \cong (G, 1).$$

If $G \subset \mathbb{Q}$, then it is the dimension group of an odometer. If $G \cap (\mathbb{R} \setminus \mathbb{Q})$ is nonempty, then there exists a Denjoy transformation φ with $D_m(X, \varphi) = G$ (see Putnam et al. (1986)). Hence we have

Corollary 3 (Giordano et al. (1995), Cor. 2.2). *Let (X, φ) be a uniquely ergodic Cantor minimal system. Then (X, φ) is orbit equivalent either to an odometer system or to a Denjoy system.*

Remark 8 The study of speedups of ergodic, finite measure-preserving transformations started in the late sixties. In Arnoux et al. (1985), they showed that if T and S are any two ergodic, measure-preserving automorphisms, there is a speedup of T that is measurably conjugate to S , or equivalently using Dye's result (1959) that if T and S are two orbit equivalent, ergodic, measure-preserving transformations, then S is measurably conjugate to a speedup of T .

In 2016, Ash studied in (2016) speedups of Cantor minimal systems and showed that if (X_i, φ_i) are two minimal Cantor systems, (X_2, φ_2) is a speedup of (X_1, φ_1) if and only if there is a unital surjective homomorphism Φ from $D_m(X_2, \varphi_2)$ onto $D_m(X_1, \varphi_1)$ such that

$$\Phi(D_m(X_2, \varphi_2)^+) = D_m(X_1, \varphi_1)^+$$

and this is in turn equivalent to the existence of a homeomorphism $F: X_1 \rightarrow X_2$ which induces an injection from $M(X_1, \mathcal{R}_{\varphi_1})$ into $M(X_2, \mathcal{R}_{\varphi_2})$.

This work was continued in a 2018 paper by Alvin, Ash, and Ormes (2018), with the additional assumption that the speedup function is bounded. They showed that a minimal bounded speedup of an odometer is conjugate to an odometer.

We now discuss the general case of actions of finitely generated abelian groups. Let φ be a free minimal action of $\mathbb{Z}^d$ on the Cantor set X and let $\mathcal{R}_\varphi$ be the associated étale equivalence relation. The affability of $\mathcal{R}_\varphi$ is proved by constructing sub-equivalence relations of $\mathcal{R}_\varphi$:

$$\mathcal{R}_0 \subset \mathcal{R}_1 \subset \dots \mathcal{R}_d = \mathcal{R}_\varphi$$

such that $\mathcal{R}_0$ is a minimal AF-equivalence relation with the relative topology from $\mathcal{R}_\varphi$, and each $\mathcal{R}_i$ is a “small” extension of $\mathcal{R}_{i-1}$ for $i = 1, \dots, d$.

Then by applying d times the absorption theorem, we get that $\mathcal{R}_i$ is orbit equivalent to $\mathcal{R}_{i-1}$, for $1 \leq i \leq d$, and therefore $\mathcal{R}_\varphi$ is orbit equivalent to $\mathcal{R}_0$, hence affable.

The details of the construction of the minimal AF subrelation $\mathcal{R}_0$ of $\mathcal{R}_\varphi$ are in Giordano et al. (2008) for $d = 2$ and in Giordano et al. (2010) for $d \geq 3$. We then have

Theorem 14 *Any free minimal action of $\mathbb{Z}^d$ on the Cantor set is affable.*

Let φ be a minimal action on the Cantor set of a finitely generated abelian group. By generalizing a result of Ø. Johansen (1998), $\mathcal{R}_\varphi$ is orbit equivalent to an action of $\mathbb{Z}^d$, with $d \geq 1$. Then by Theorem 1.46 we get

Theorem 15 *Any minimal action of a finitely generated abelian group on the Cantor set is affable.*

As a consequence of Theorems 13 and 15, we then have

Theorem 16 *Let $(X, \mathcal{R})$ and $(X', \mathcal{R}')$ be two minimal equivalence relations on Cantor sets which are either AF-relations or arise from actions of finitely generated abelian groups.*

Then they are orbit equivalent if and only if

$$(D_m(X, \mathcal{R}), D_m(X, \mathcal{R})^+, [\mathbf{1}_X]) \\ \cong (D_m(X', \mathcal{R}'), D_m(X', \mathcal{R}')^+, [\mathbf{1}_{X'}]).$$

Then as a corollary of Theorem 8, we have

Corollary 4 *For any simple dimension groups G with a trivial infinitesimal subgroup, there exists a Cantor minimal system (X, φ) such that $D_m(X, \varphi) = G$.*

Remark 9 Let $d \geq 2$. By Theorem 16 and Corollary 4, for any free minimal action of $\mathbb{Z}^d$, $D_m(X, \mathcal{R}_{\mathbb{Z}^d})$ is a simple, acyclic dimension group with no nontrivial infinitesimal elements, but an exact description of the range of the invariant D_m is not yet known.

However, using results of M.-I. Cortez and S. Petite (2014), S.-M. Høynes (2016) showed that it contains all unital simple dimension groups (G, u) , with no nontrivial infinitesimal elements, and a noncyclic rational subgroup $\mathbb{Q}(G, u)$.

A Topological Krieger Theorem: The Notion of Strong Orbit Equivalence

In this subsection, (X, φ) will always denote a Cantor minimal system. The motivation of Giordano et al. (1995) was to obtain a topological version of the famous Krieger theorem.

Theorem 17 (Krieger 1976). *Let (X_1, μ_1, T_1) and (X_2, μ_2, T_2) be two ergodic nonsingular systems and $M_1 = W^*(X_1, \mu_1, T_1)$ and $M_2 = W^*(X_2, \mu_2, T_2)$ their associated von Neumann–Krieger factors. Then the following are equivalent:*

1. *The dynamical systems are orbit equivalent, that is, there exists a bimeasurable bijection $F: X_1 \rightarrow X_2$ preserving the measure class, such that, for almost all $x \in X_1$, $F(\text{Orb}_{T_1}(x)) = \text{Orb}_{T_2}(F(x))$.*
2. *The von Neumann–Krieger factors M_1 and M_2 are isomorphic.*
3. *the Krieger associated flow of (X_1, μ_1, T_1) and of (X_2, μ_2, T_2) are conjugate or equivalently the flow of weights of M_1 and of M_2 are isomorphic.*

Let (X, φ) be a Cantor minimal system and $C^*(X, \varphi)$ be the associated C^* -crossed product $C(X) \rtimes_{\varphi} \mathbb{Z}$. Recall (see, e.g., Giordano et al. (1995), p. 60 for the precise references) that

1. $C^*(X, \varphi)$ is a simple, unital, AT-algebra of real rank zero and stable rank one, hence is classified by its Elliott invariant;
2. If $K_i(C^*(X, \varphi))$ denotes, for $i = 0, 1$, the K -groups of the crossed product, then $K_0(C^*(X, \varphi))$ is a simple dimension group order isomorphic to $D(X, \varphi)$, and $K_1(C^*(X, \varphi)) = \mathbb{Z}$ (see [?]).

Hence, we have

Theorem 18 *A complete isomorphism invariant for the family of $C^*(X, \varphi)$, where (X, φ) is a Cantor minimal system, is the simple dimension group $(D(X, \varphi), D(X, \varphi)^+, [\mathbf{1}_X])$.*

Remark 10

- (1) Boyle and Handelman (1994) showed that all possible (topological) entropies can be realized in the class of Cantor minimal system (X, φ) , such that $(D(X, \varphi), [\mathbf{1}])$ is order isomorphic to $(\mathbb{Z}[\frac{1}{2}], 1)$. Therefore,

$$C^*(X_1, \varphi_1) \cong C^*(X_2, \varphi_2) \\ \Leftrightarrow (X_1, \varphi_1) \text{ conjugate to } (X_2, \varphi_2).$$

Sugisaki (1998, 2003) generalized Boyle–Handelman’s result to the class of Cantor minimal system (X, φ) such that $(D(X, \varphi), D(X, \varphi)^+, [\mathbf{1}_X])$ is a fixed simple dimension group (G, G^+, u) .

- (2) Let (X, φ) be a Cantor minimal system. By Proposition 6 and Theorem 13, the unital simple dimension group $(D(X, \varphi)/\text{Inf}(D(X, \varphi)), [\mathbf{1}_X])$ is a complete invariant of orbit equivalence.

Consequently, the orbit equivalence of two Cantor minimal systems does not imply in general the isomorphism of the corresponding crossed products.

Let (X_1, φ_1) and (X_2, φ_2) be two orbit equivalent Cantor minimal systems and $F : X_1 \rightarrow X_2$ be an orbit equivalence. Then F defines two maps $m, n : X_1 \rightarrow \mathbb{Z}$ given for $x \in X$, by

$$\begin{aligned} F(\varphi_1(x)) &= \varphi_2^{n(x)}(F(x)) \text{ and } F\left(\varphi_1^{m(x)}(x)\right) \\ &= \varphi_2(F(x)). \end{aligned}$$

We will call n and m the *orbit cocycles* associated to F .

Definition 8 Two Cantor minimal systems (X_1, φ_1) and (X_2, φ_2) are strongly orbit equivalent (SOE) if there exists an orbit equivalence $F : X_1 \rightarrow X_2$ such that each of the associated orbit cocycles $m, n : X_1 \rightarrow \mathbb{Z}$ has at most one point of discontinuity.

Strong orbit equivalence being an equivalence relation will follow from the next theorem.

Theorem 19 (Giordano et al. (1995), Theorem 2.1). *Let (X_1, φ_1) and (X_2, φ_2) be two orbit equivalent Cantor minimal systems. The following statements are equivalent:*

- (1) (X_1, φ_1) and (X_2, φ_2) are strongly orbit equivalent.
- (2) $(D(X_1, \varphi_1), D(X_1, \varphi_1)^+, [\mathbf{1}_{X_1}])$ and $(D(X_2, \varphi_2), D(X_2, \varphi_2)^+, [\mathbf{1}_{X_2}])$ are order isomorphic.
- (3) $C^*(X_1, \varphi_1) \cong C^*(X_2, \varphi_2)$.

Note that

- a) The equivalence between the statements (2) and (3) follows from Theorem 18
- b) If (X, φ_1) and (X, φ_2) are strongly orbit equivalent, then the coboundary subgroups $B(X, \varphi_1)$ and $B(X, \varphi_2)$ are equal and (2) follows

Remark 11 Recall that the famous Jewett–Krieger theorem (see, e.g., Glasner (2003)) states that any ergodic transformation of a nonatomic Lebesgue space has a uniquely ergodic minimal Cantor model.

Ormes (1997) presents Jewett–Krieger-type realization results of an ergodic dynamical system

by a Cantor minimal system in a prescribed orbit equivalence class. More precisely, he proves in (Ormes 1997, Theorem 7.2):

Theorem 20 *Let (Y, ν, T) be an ergodic probability measure-preserving dynamical system, (X, φ) a Cantor minimal system, and $\mu \in M(X, \varphi)$ an ergodic measure.*

Then there is a Cantor minimal system (X, ψ) , orbit equivalent to (X, φ) such that (Y, ν, T) is measurably conjugate to (X, ψ, μ) .

Moreover in Ormes (1997, Theorem 27), he gives necessary and sufficient spectral conditions for a Jewett–Krieger-type realization of an ergodic dynamical system by a Cantor minimal system in a prescribed strong orbit equivalence class.

Continuous Orbit Equivalence

In this section, a topological dynamical system (X, G) will denote a countable discrete group G acting on a compact, second countable, Hausdorff space X . Recall that (X, G) is topologically free if for every $g \in G, g \neq e$, the set $\{x \in X \mid gx \neq x\}$ is dense in X or equivalently if the subset of elements $x \in X$ for which the stabilizer subgroup $G_x = \{g \in G \mid gx = x\}$ is trivial is dense in X .

Definition 9 Two topological dynamical systems (X_1, G_1) and (X_2, G_2) are continuously orbit equivalent if there exist a homeomorphism $F : X_1 \rightarrow X_2$ and two continuous maps $\alpha_1 : X_1 \times G_1 \rightarrow G_2$ and $\alpha_2 : X_2 \times G_2 \rightarrow G_1$ such that

- (1) $F(g_1 x_1) = \alpha_1(x_1, g_1) F(x_1), x_1 \in X_1, g_1 \in G_1$
- (2) $F^{-1}(g_2 x_2) = \alpha_2(x_2, g_2) F^{-1}(x_2), x_2 \in X_2, g_2 \in G_2$

Remark 12 If for $i = 1$ and 2 , (X_i, G_i) is topologically free, then α_1 and α_2 are uniquely determined by the conditions (1) and (2) of Definition 9 and are continuous one-cocycles (see, e.g., Li (2018a)).

For two topological dynamical systems (X_i, G_i) , let us recall that they are *isomorphic* if there is a homeomorphism $F : X_1 \rightarrow X_2$ and a

group isomorphism $\gamma : G_1 \rightarrow G_2$ such that for all $g \in G_1, x \in X_1$,

$$F(gx) = \gamma(g)F(x).$$

The notion of continuous orbit equivalence was initially studied by M. Boyle in his thesis (Boyle 1983), where he proved:

Proposition 9 *For $i = 1, 2$, let φ_i be a topologically transitive homeomorphism of a compact metrizable space X_i . If φ_1 and φ_2 are continuously orbit equivalent, then φ_1 and φ_2 are flip conjugate, that is, φ_1 is conjugate to φ_2 or to φ_2^{-1} or the $\mathbb{Z}$ -dynamical systems (X_1, φ_1) and (X_2, φ_2) are isomorphic.*

This proposition is the analogue in the ergodic probability measure-preserving case to Belinskaya's result: the integrability of one of the associated orbit cocycles implies flip conjugacy (in the measurable category).

Boyle and Tomiyama (1998) generalized this result as follows:

Theorem 21 *For $i = 1, 2$, let φ_i be a topologically free homeomorphism of a compact Hausdorff space X_i . If φ_1 and φ_2 are continuously orbit equivalent, then there exist open partitions $X_1 = A_1 \amalg A_2$ and $X_2 = B_1 \amalg B_2$ such that $\varphi_1|_{A_1}$ is conjugate to $\varphi_2|_{B_1}$ and $\varphi_1|_{A_2}$ is conjugate to $\varphi_2^{-1}|_{B_2}$.*

Remark 13

- 1) In measurable dynamics, orbit equivalence rigidity is an important domain of research and several impressive results have been proved (see, e.g., Furman (1999); Ioana (2011); Monod and Shalom (2006); Popa (2007)). In topological dynamics, M. Boyle's result was the first proved positive rigidity result. Li (2018a) obtains both positive and negative rigidity results.
- 2) In Li (2018a), the author defines a topologically free dynamical system (X, G) to be *continuous orbit equivalence rigid* if any topologically free dynamical system

continuous orbit equivalent to (X, G) is isomorphic to it. Then he gives several examples of topologically free dynamical systems which are continuous orbit equivalence rigid.

- 3) Medynets, Sauer and Thom, in their study of Cantor systems and quasi-isometry of groups (Medynets et al. 2017), prove that two finitely generated non-amenable groups are quasi-isometric if and only if they admit continuous orbit equivalent Cantor minimal actions. In particular, free groups of different rank admit continuous orbit equivalent Cantor minimal actions, unlike in the measurable setting.

Li (2018b), by studying a weaker form of continuous orbit equivalence, extends Medynets, Sauer, and Thom's result and proves in particular that two finitely generated groups are quasi-isometric if and only if there exist Kakutani equivalent topologically free systems of these groups on compact spaces.

Continuous Orbit Equivalence and C^* -algebra

Let (X, G) be a topological dynamical system and $X \rtimes G$ be the associated topological transformation groupoid. Then it is an étale groupoid and is a principal groupoid (i.e., $X \rtimes G = \mathcal{R}_G$) if the action is free, and $X \rtimes G$ is essentially principal if the action is topologically free. Recall that the reduced C^* -algebra $C_r^*(X \rtimes G)$ of the transformation groupoid $X \rtimes G$ is canonically isomorphic to the C^* -crossed product $C_0(X) \rtimes G$ (see, e.g., Li (2018a) or Matui (2017)).

Remark 14 If (X, G) is topologically free, then the pair $(C_0(X) \rtimes G, C_0(X))$ is a Cartan pair as defined by J. Renault (2008).

Generalizing a classical result of I. Singer (1955) on the interplay between measured dynamical systems and the Murray–von Neumann group measure space construction, X. Li (2018a) proves:

Theorem 22 *Let (X_i, G_i) be two topologically free dynamical systems as above. Then the following are equivalent:*

- (1) (X_1, G_1) and (X_2, G_2) are continuously orbit equivalent.
- (2) The transformation groupoids $X_1 \rtimes G_1$ and $X_2 \rtimes G_2$ are isomorphic.
- (3) There is a C^* -isomorphism Φ from $C_0(X_1) \rtimes G_1$ onto $C_0(X_2) \rtimes G_2$ such that $\Phi(C_0(X_1)) = C_0(X_2)$.

Remark 15

- 1) The equivalence of (2) and (3) is due to J. Renault (2008).
- 2) J. Tomiyama proved the equivalence of (1) and (3) for topologically free $\mathbb{Z}$ -actions in Tomiyama (1996).
- 3) Theorem 22 is extended to more general groupoids in a recent publication of Carlsen, Ruiz, Sims, and Tomforde (2021).

For $\mathbb{Z}$ -actions, using Theorem 23, M. Boyle and J. Tomiyama (1998) proved the following rigidity result:

Theorem 23 For $i = 1, 2$, let φ_i be a topologically free homeomorphism of a compact Hausdorff space X_i . Then the following are equivalent:

- (1) There is a C^* -isomorphism $\Phi : C(X_1) \rtimes_{\varphi_1} \mathbb{Z} \rightarrow C(X_2) \rtimes_{\varphi_2} \mathbb{Z}$ such that $\Phi(C(X_1)) = C(X_2)$.
- (2) There exist open partitions $X_1 = A_1 \coprod A_2$ and $X_2 = B_1 \coprod B_2$ such that $\varphi_1|_{A_1}$ is conjugate to $\varphi_2|_{B_1}$ and $\varphi_1|_{A_2}$ is conjugate to $\varphi_2^{-1}|_{B_2}$.

If φ_1 and φ_2 are topologically transitive or the spaces are connected, then these conditions hold if and only if φ_1 and φ_2 are flip conjugate.

Remark 16 Theorem 21 was proved for Cantor minimal systems in Giordano et al. (1995) using Proposition 9.

$\mathbb{Z}^d$ -odometers

We have already introduced the notions of isomorphism, continuous orbit equivalence, and (topological) orbit equivalence between two topological dynamical systems (X_1, G_1) and (X_2, G_2) . Before adding a strengthening of isomorphism, let us denote by φ_i the action of G_i on X_i , for $i = 1, 2$.

If $G_1 = G_2 = G$, and if there exists a homeomorphism $F : X_1 \rightarrow X_2$ such that $F \circ \varphi_1(g) = \varphi_2(g) \circ F$ for all $g \in G$, we will say that (X_1, φ_1) and (X_2, φ_2) are conjugate.

For two topological dynamical systems, the following implications are clear: conjugacy $\Rightarrow$ isomorphism $\Rightarrow$ continuous orbit equivalence $\Rightarrow$ orbit equivalence.

In this section, we characterize (algebraically) these four types of equivalences for $\mathbb{Z}$ - and $\mathbb{Z}^d$ -odometers. Let us first recall the definition of $\mathbb{Z}^d$ -odometers.

Let $\mathcal{G}$ be a (nontrivial) decreasing sequence $(G_n)_{n \geq 1}$ of finite index subgroups of $\mathbb{Z}^d$. For $n \geq 1$, let $q_n : \mathbb{Z}^d/G_{n+1} \rightarrow \mathbb{Z}^d/G_n$ be the quotient map and let $X_{\mathcal{G}}$ be the corresponding inverse limit. Let $\phi_{\mathcal{G}}$ be the profinite action of $\mathbb{Z}^d$ on the $X_{\mathcal{G}}$ given by

$$\phi_{\mathcal{G}}(v)(x_1, x_2, \dots) = (x_1 + v, x_2 + v, \dots) \quad (v \in \mathbb{Z}^d).$$

Then following Cortez (2006), a $\mathbb{Z}^d$ -odometer is any $\mathbb{Z}^d$ -Cantor minimal system conjugate to $(X_{\mathcal{G}}, \phi_{\mathcal{G}})$ for some $\mathcal{G}$ as above. Note that

- (1) $(X_{\mathcal{G}}, \phi_{\mathcal{G}})$ is free if and only if $\bigcap_{n=1}^{\infty} G_n = \{0\}$
- (2) The free $\mathbb{Z}^d$ -Cantor minimal system $(X_{\mathcal{G}}, \phi_{\mathcal{G}})$ is equicontinuous

Remark 17 G -odometers can be defined for any residually finite group G (see Cortez and Petite (2014) or Downarowicz (2005)). Moreover in Cortez and Medynets (2016), the authors show that if a dynamical system (X, H) is a free minimal equicontinuous system, then H is necessarily residually finite and (X, H) is conjugate to a H -odometer.

Following Giordano et al. (2019), let us recall another presentation of $\mathbb{Z}^d$ -odometers using Pontryagin duality. Let H be a subgroup of $\mathbb{Q}^d$ containing $\mathbb{Z}^d$. Then we have

$$H/\mathbb{Z}^d \subset \mathbb{Q}^d/\mathbb{Z}^d \subset \mathbb{R}^d/\mathbb{Z}^d \cong \mathbb{T}^d.$$

Let ρ denote the canonical inclusion of $H/\mathbb{Z}^d$ in $\mathbb{T}^d$. By duality, let $\hat{\rho}$ be the homomorphism from

$\mathbb{Z}^d$ to Y_H , the Pontryagin dual $\widehat{H/\mathbb{Z}^d}$ of $H/\mathbb{Z}^d$, which is a compact, totally disconnected space.

Then let (Y_H, ψ_H) be the $\mathbb{Z}^d$ -dynamical system where ψ_H denotes the action of $\mathbb{Z}^d$ on Y_H given by

$$\psi_H^n(x) = x + \widehat{\rho}(n), \text{ for } n \in \mathbb{Z}^d, x \in Y_H.$$

Note that

- (1) If $\mathbb{Z}^d \subset H \subset \mathbb{Q}^d$, then (Y_H, ψ_H) is free if and only if H is dense in $\mathbb{Q}^d$.
- (2) If $\mathbb{Z}^d \subset H_1 \subset H_2 \subset \mathbb{Q}^d$, then there is a natural factor map from (Y_{H_2}, ψ_{H_2}) to (Y_{H_1}, ψ_{H_1}) .
- (3) If $\mathbb{Z}^d \subset H_1 \subset H_2 \subset \cdots \subset \mathbb{Q}^d$, then the inverse limit of the $\mathbb{Z}^d$ -dynamical systems $(Y_{H_n}, \psi_{H_n})_{n \geq 1}$ is conjugate to (Y_H, ψ_H) , where $H = \bigcup_{n \geq 1} H_n$.

To describe the link between the two constructions, recall that a finite-index subgroup G of $\mathbb{Z}^d$ is a lattice of $\mathbb{R}^d$, and that its dual lattice G^* is a finite-index extension of $\mathbb{Z}^d$.

Then (see Giordano et al. (2019), Theorems 27 and 28) one has that

- a) The $\mathbb{Z}^d$ -odometer (X_G, ϕ_G) , given by a (nontrivial) decreasing sequence $G = (G_n)_{n \geq 1}$ of finite index subgroups of $\mathbb{Z}^d$ is conjugate to (Y_H, ψ_H) , where $H = \bigcup_{n \geq 1} G_n^*$, and conversely
- b) If $(H_n)_{n \geq 1}$ is an increasing sequence of finite index extensions of $\mathbb{Z}^d$ and $H = \bigcup_{n \geq 1} H_n \subset \mathbb{Q}^d$, then $G = (H_n^*)_{n \geq 1}$ is a decreasing sequence of finite index subgroups of $\mathbb{Z}^d$, and (Y_H, ψ_H) and (X_G, ϕ_G) are conjugate.

Before stating the main result of Giordano et al. (2019), let us recall (Giordano et al. (2019), Def. 1.4) that if H is a subgroup $\mathbb{Z}^d \subset H \subset \mathbb{Q}^d$, the *superindex* of $\mathbb{Z}^d$ in H is defined as

$$[[H : \mathbb{Z}^d]] = \left\{ [H' : \mathbb{Z}^d] \mid \mathbb{Z}^d \subset H' \subset H, [H' : \mathbb{Z}^d] < \infty \right\}.$$

Theorem 24 (Giordano et al. (2019), Theorem 1.5). *Let $\mathbb{Z}^d \subset H \subset \mathbb{Q}^d$ and $\mathbb{Z}^{d'} \subset H' \subset \mathbb{Q}^{d'}$ be two dense subgroups of $\mathbb{Q}^d$ and, respectively, $\mathbb{Q}^{d'}$.*

- (1) *If $d, d' \leq 2$, then (Y_H, ψ_H) and $(Y_{H'}, \psi_{H'})$ are conjugate if and only if $d = d'$ and $H = H'$.*
- (2) *If $d, d' \leq 2$, then (Y_H, ψ_H) and $(Y_{H'}, \psi_{H'})$ are isomorphic if and only if $d = d'$ and there exists $\alpha \in GL_d(\mathbb{Q})$ with $\det(\alpha) = \pm 1$ such that $\alpha H = H'$.*
- (3) *If $d, d' \leq 2$, then (Y_H, ψ_H) and $(Y_{H'}, \psi_{H'})$ are continuously orbit equivalent if and only if $d = d'$ and there exists $\alpha \in GL_d(\mathbb{Z})$ such that $\alpha H = H'$.*
- (4) *The $\mathbb{Z}^d$ -action (Y_H, ψ_H) and the $\mathbb{Z}^{d'}$ -action $(Y_{H'}, \psi_{H'})$ are orbit equivalent if and only if the superindex $[[H : \mathbb{Z}^d]]$ of $\mathbb{Z}^d$ in H is equal to superindex $[[H' : \mathbb{Z}^{d'}]]$.*

Remark 18

- 1) As any two orbit equivalent $\mathbb{Z}$ -odometers are conjugate (see, e.g., Giordano et al. (2019), Cor. 5.9), the four conditions of Theorem 24 are equivalent for $d = d' = 1$.

The situation is remarkably different for $\mathbb{Z}^2$ -odometers as the four conditions are all inequivalent (see Cortez (2006) and Li (2018a) or Giordano et al. (2019), Example 5.10).

- 2) Cortez and Medynets (2016) study the notion of continuous orbit equivalence for G -odometers, where G is a residually finite group. They study in particular the topological full group associated to a free G -odometer.
- 3) We reviewed in Remark 8 the research in topological speedups of Cantor minimal systems. Johnson and McClendon (2022) study these topological notions in the context of actions of $\mathbb{Z}^d$, and in particular for $\mathbb{Z}^d$ -odometers.

Cohomology of $\mathbb{Z}^d$ -odometers

To any dense subgroup $\mathbb{Z}^d \subset H \subset \mathbb{Q}^d$, we have associated a $\mathbb{Z}^d$ -action (Y_H, ψ_H) on the Cantor set Y_H . In this subsection, we recall the definition of the integer-valued cohomology of an action of $\mathbb{Z}^d$ on the Cantor set and show that $H^1(Y_H, \psi_H) \cong H$.

Let (X, φ) denote a $\mathbb{Z}^d$ -action on the Cantor set X . Recall that $C(X, \mathbb{Z})$, the abelian group of continuous $\mathbb{Z}$ -valued functions on X , is also a $\mathbb{Z}^d$ -module (for $n \in \mathbb{Z}^d$ and $f \in C(X, \mathbb{Z})$, $n \cdot f(x) = f(\varphi(n)x)$, $x \in X$).

The integer-valued cohomology of (X, φ) can be defined as the group cohomology of $\mathbb{Z}^d$ with coefficients in the module $C(X, \mathbb{Z})$, or as the Čech cohomology of the mapping torus associated to (X, φ) (see, e.g., Forrest and Hunton (1999), Hunton (2015)).

Remark 19 If Γ is a discrete countable group and (X, φ) is a Γ -action on the Cantor set, let $\mathcal{G}_\varphi$ denote the étale transformation groupoid. Recall that $\mathcal{G}_\varphi = \Gamma^d \times X$ and the groupoid structure is defined as follows: (γ, x) and (γ', y) are composable if and only if $x = \varphi(\gamma')(y)$, and

$$(\gamma, \varphi(\gamma')(y))(\gamma', y) = (\gamma \cdot \gamma', y) \text{ and } (\gamma, x)^{-1} = (\gamma^{-1}, \varphi(\gamma)(x)).$$

Then, the cohomology groups $H^*(\Gamma, C(X, \mathbb{Z}))$ are canonically isomorphic to the cohomology groups $H^*(\mathcal{G}_\varphi)$ associated to the étale groupoid $\mathcal{G}_\varphi$. The cohomology $H^*(\mathcal{G})$ for any étale groupoid $\mathcal{G}$ were introduced by J. Renault (1980) and has been widely studied.

Let (X, φ) be a free minimal $\mathbb{Z}^d$ -Cantor system. Then the following holds:

Lemma 1

(1) $H^d(X, \varphi)$ is the group of co-invariants, that is,

$$H^d(X, \varphi) = D(X, \varphi).$$

- (2) $H^0(X, \varphi) = \{f \in C(X, \mathbb{Z}) \mid f \circ \varphi(n) = f, \text{ for all } n \in \mathbb{Z}^d\} = \mathbb{Z}$
- (3) By Poincaré duality, there is a natural isomorphism between $H^n(X, \varphi)$ and the homology group $H_{d-n}(X, \varphi) := H_{d-n}(\mathbb{Z}; C(X, \mathbb{Z}))$.

and as continuous orbit equivalence of (X, φ) corresponds to isomorphism of the transformation groupoid, we get

Proposition 10 The cohomology $H^*(\mathcal{G}_\varphi)$ is an invariant of continuous orbit equivalence.

Remark 20 The homology groups $H_*(\mathcal{G})$ of an étale groupoid were first introduced in Crainic and Moerdijk (2000). When the unit space of the groupoid $\mathcal{G}$ is totally disconnected, H. Matui (2012, 2015, 2016) explored the connections between the homology and the dynamical properties of $\mathcal{G}$. He studied in particular the interplay between the homology groups, the topological full group, and also the K -theory of $C_r^*(\mathcal{G})$, the reduced groupoid C^* -algebra of $\mathcal{G}$ (see, e.g., Matui's review paper (2017)).

As above, let (X, φ) be a free minimal $\mathbb{Z}^d$ -Cantor system. Let us first consider the first group of cohomology of (X, φ) . Recall that a continuous function $\theta : X \times \mathbb{Z}^d \rightarrow \mathbb{Z}$ is a 1-cocycle if and only if

$$\theta(x, m+n) = \theta(x, m) + \theta(\varphi(m)(x), n) \text{ for all } x \in X \text{ and } m, n \in \mathbb{Z}^d.$$

A cocycle θ is a coboundary if there is $h \in C(X, \mathbb{Z})$ such that $\theta(x, m) = h(\varphi(m)(x)) - h(x)$ for all $x \in X$ and $m \in \mathbb{Z}^d$.

Let (Y_H, ψ_H) be the $\mathbb{Z}^d$ -odometer associated to a dense subgroup $\mathbb{Z}^d \subset H \subset \mathbb{Q}^d$. Then using Giordano et al. (2019), Theorem 2.5, and a simple case of Shapiro's Lemma, we have

Theorem 25 (Giordano et al. (2019), Theorem 4.3). Let $\mathbb{Z}^d \subset H \subset \mathbb{Q}^d$. Then $H^1(Y_H, \psi_H) \cong H$.

For $1 \leq d \leq 2$, a precise description can be given for such an isomorphism. Note first that if μ is an invariant probability measure of a $\mathbb{Z}^d$ -Cantor system (X, φ) and θ is a 1-cocycle of (X, φ) , then the map $\tau_\mu^1(\theta) : \mathbb{Z}^d \rightarrow \mathbb{R}$ defined for $n \in \mathbb{Z}^d$ by

$$\tau_\mu^1(\theta)(n) = \int_X \theta(x, n) d\mu(x),$$

is a group homomorphism. As τ_μ^1 vanishes on coboundaries, it defines a group homomorphism $\tau_\mu^1 : H^1(X, \varphi) \rightarrow \text{Hom}(\mathbb{Z}^d, \mathbb{R})$.

Using the canonical basis $(e_n)_{1 \leq n \leq d}$ of $\mathbb{Z}^d$ to identify $\text{Hom}(\mathbb{Z}^d, \mathbb{R})$ with $\mathbb{R}^d$, we define $\tau_\mu^1 : H^1(X, \varphi) \rightarrow \mathbb{R}$ for any 1-cocycle θ by

$$\tau_\mu^1([\theta]) = \left(\tau_\mu^1(\theta)(e_1), \dots, \tau_\mu^1(\theta)(e_d) \right).$$

We then get

Theorem 26 (Giordano et al. (2019), Theorem 4.4). *Let $1 \leq d \leq 2$. (Y_H, ψ_H) be the $\mathbb{Z}^d$ -odometer associated to a dense subgroup $\mathbb{Z}^d \subset H \subset \mathbb{Q}^d$, and μ be its unique invariant probability measure. Then*

$$\tau_\mu^1 : H^1(X, \varphi) \rightarrow H$$

is an isomorphism.

Mean Dimension, Small Boundary Property, and Classification of C^* -Algebras

The theory of mean dimension for topological dynamical systems was suggested by Gromov and largely developed by Lindenstrauss and Weiss (2000). In many ways, it parallels the theory of covering dimension for topological spaces. The small boundary property, on the other hand, can be viewed as a dynamical version of total disconnectedness (see Definition 20). These two closely related notions have profound implications to several long-standing open questions in topological dynamics (see, e.g., Gutman et al. (2016); Lindenstrauss (2000); Lindenstrauss and Weiss (2000)).

In the last decade, these purely dynamical notions have found deep connections with the classification of simple nuclear C^* -algebras, with the central theme being the study of “regularity” for crossed products. As we will see, ideas and techniques that originate from the Kakutani–Rokhlin tower decomposition, or more generally, the Ornstein–Weiss tiling theorem in ergodic theory (Ornstein and Weiss 1987), are fundamental in many of these connections. Interestingly, the attempt to understand the structure of crossed products also led to the discovery of some purely dynamical results, such as a new characterization of the small boundary property.

We use the following notation and terminologies throughout this chapter. Let G be a discrete group and X be a compact Hausdorff space. A G -action is a group homomorphism α from G into the homeomorphism group of X . In this case, we also write $\alpha : G \curvearrowright X$ for the action and say that X is a G -space.

Given a G -action $\alpha : G \curvearrowright X$, there is an induced action of G on the C^* -algebra $C(X)$ of continuous functions on X given by

$$s.f(x) := f(s^{-1}x)$$

for $s \in G$ and $f \in C(X)$, $x \in X$. We write $C(X) \rtimes_r G$ for the reduced crossed product C^* -algebras. If we need to specify the action, then we write $C(X) \rtimes_{\alpha,r} G$ and $C(X) \rtimes_{\alpha,r} G$.

Example 6 Let $G \curvearrowright X$ be a free minimal action of a countable discrete amenable group G on a compact Hausdorff space X . The reduced crossed product $C(X) \rtimes_r G$ can be realized in the following way: let μ be a G -invariant probability measure on X (which exists because G is amenable) and H be the Hilbert space $L^2(X, \mu) \otimes \ell^2(G)$. We define a representation $\pi : C(X) \rightarrow B(H)$ by

$$\pi(f)(\varphi \otimes \delta_s) := (s^{-1}.f)\varphi \otimes \delta_s$$

for $f \in C(X)$, $\varphi \in L^2(X, \mu)$ and $s \in G$. Let $\lambda : G \rightarrow \mathcal{U}(\ell^2(G))$ denote the left regular representation, that is, $\lambda_g(\delta_s) = \delta_{gs}$. The reduced crossed product $C(X) \rtimes_r G$ is the C^* -algebra generated by the images $\pi(C(X))$ and $1 \otimes \lambda(G)$ in $B(H)$.

We refer the readers to, for example, (Brown and Ozawa (2008), Chap. 4) for the general constructions and basic properties of these algebras.

Mean Dimension and Radius of Comparison

We start by recalling the definition of mean dimension for a single homeomorphism from Lindenstrauss and Weiss (2000). Let X be a compact metrizable space and $T : X \rightarrow X$ be a homeomorphism. For two open covers α and β of X , we write $\beta \succ \alpha$ (or $\alpha < \beta$) if β is an open refinement of α , and let $\alpha \vee \beta$ denote the join of α and β . Define the *order* of an open cover α by

$$\text{ord}(\alpha) := \max_{x \in X} \left(\sum_{U \in \alpha} 1_U(x) - 1 \right),$$

where 1_U is the characteristic function of U , and set

$$\mathcal{D}(\alpha) := \min\{\text{ord}(\beta) : \beta \succ \alpha\}.$$

Definition 10 Let X be a compact metrizable space and T be a homeomorphism T of X . Its mean dimension $\text{mdim}(X, T)$ is given by

$$\sup_{\alpha} \left(\lim_{n \rightarrow \infty} \frac{\mathcal{D}(\alpha \vee T^{-1}\alpha \vee \dots \vee T^{-(n-1)}\alpha)}{n} \right),$$

where the supremum is taken over all open covers α of X .

This definition should be compared to the Lebesgue covering dimension (see, e.g., Pears (1975), Chap. 3), which is given by

$$\dim(X) := \sup_{\alpha} D(\alpha).$$

Although we only gave the definition in the case of a single homeomorphism, the definition can be generalized to actions of countable amenable groups (e.g., see Coornaert (2015)) and to actions of the even larger class of countable sofic groups (Li 2013).

Example 7 If X has finite Lebesgue covering dimension, then the mean dimension of any homeomorphism T is necessarily zero. Moreover, by Lindenstrauss and Weiss (2000, Theorem 5.4) every system (X, T) with finite topological entropy has mean dimension zero.

Example 8 Let K be a compact metric space of finite covering dimension and consider the shift homeomorphism $\sigma : K^{\mathbb{Z}} \rightarrow K^{\mathbb{Z}}$. By Lindenstrauss and Weiss (2000, Proposition 3.1) we have

$$\text{mdim}(K^{\mathbb{Z}}, \sigma) \leq \dim(K).$$

Moreover, if $K = [0, 1]^d$ for some positive integer d , then the equality holds (see Lindenstrauss and Weiss (2000), Proposition 3.3), that is,

$$\text{mdim}\left(\left([0, 1]^d\right)^{\mathbb{Z}}, \sigma\right) = \dim([0, 1]^d) = d.$$

It was a long-standing open problem whether every minimal system can be embedded into the two-sided shift $([0, 1]^{\mathbb{Z}}, \sigma)$. It is not hard to see that the mean dimension cannot increase when passing to closed invariant subsets, so for a dynamical system (X, T) to be embeddable into the shift $(([0, 1]^d)^{\mathbb{Z}}, \sigma)$, it is necessary to have $\text{mdim}(X, T) \leq d$. Lindenstrauss and Weiss (2000) constructed a minimal subshift of $(([0, 1]^2)^{\mathbb{Z}}, \sigma)$ with mean dimension strictly larger than one, thereby resolving the question in the negative.

Theorem 27 (Lindenstrauss and Weiss (2000), Proposition 3.5). *There is a minimal invariant closed subset X of $([0, 1]^2)^{\mathbb{Z}}, \sigma$ such that $\text{mdim}(X, \sigma|_X) > 1$.*

Using a similar technique Giol and Kerr (2010) constructed a minimal subshift (X, T) whose crossed product C^* -algebras $C(X) \rtimes_r \mathbb{Z}$ have non-zero radius of comparison (see Definition 12) and thereby gave the first example of a simple crossed product of the form $C(X) \rtimes_r \mathbb{Z}$ that does not fall within Elliott's classification program (see section "Classification of Simple Nuclear C^* -Algebras" for more details). In some sense, both the mean dimension and the radius of comparison are invariants that are designed for highly complicated structure.

The relationship between these two notions seems to go far beyond the apparent analogy. The main objective of this section is to introduce a conjecture that relates these two invariants in a definitive, quantitative way (Conjecture 1).

In what follows, we briefly recall the definition of the radius of comparison and fix our notations. For a C^* -algebra A , let $A_+ = \{a^*a : a \in A\}$ be the set of positive elements in A .

The following notion of subequivalence was introduced by Cuntz (1978), which is reminiscent of the Murray–von Neumann subequivalence on projections in a von Neumann algebra. It can be shown that $\sim$ is an equivalence relation on A_+ (see, e.g., Giordano et al. (2018), Proposition

2.0.1)). We refer the reader to Giordano et al. (2018, Part I) and Ara et al. (2011) for more on the Cuntz subequivalence relation.

Definition 11 For two positive elements a, b in A_+ , we say a is Cuntz subequivalent to b , denoted $a \precsim b$, if there exists a sequence $(r_n)_n$ in A such that $\lim_{n \rightarrow \infty} \|r_n b r_n^* - a\| = 0$. If $a \precsim b$ and $b \precsim a$, then we write $a \sim b$.

Example 9 Let A be the algebra $M_n(\mathbb{C})$ of $n \times n$ matrices. For two positive elements a and b in A , we have $a \precsim b$ if and only if $\text{rank}(a) \leq \text{rank}(b)$.

Example 10 Let X be a compact Hausdorff space and let $f, g \in C(X)_+$. Then $f \precsim g$ if and only if $\text{supp}(f) \subseteq \text{supp}(g)$, where $\text{supp}(f) := f^{-1}(0, +\infty)$ (see, e.g., Ara et al. (2011), p. 6, Proposition 2.5).

Example 11 Let $G \curvearrowright X$ be an action of a countable discrete group G on a Cantor space X , and let A and B be two clopen sets. Suppose there is a clopen partition $A = C_1 \sqcup C_2 \cdots \sqcup C_n$ and group elements $s_1, s_2, \dots, s_n$ such that the sets $s_1 C_1, s_2 C_2, \dots, s_n C_n$ are pairwise disjoint subsets of B , then the characteristic function χ_A is Cuntz subequivalent to the characteristic function χ_B in the crossed product $C(X) \rtimes_r G$.

To avoid technicalities, from now on we only consider the class of unital stably finite exact C^* -algebras. It includes all the crossed products that we are interested in (see, e.g., Blackadar (2006) for relevant C^* -algebraic definitions).

Recall that a *tracial state* on a C^* -algebra A is a state (i.e., a positive linear functional with norm one) φ that satisfies $\varphi(ab) = \varphi(ba)$ for all a and b in A . Let $T(A)$ denote the set of tracial states on A . By combining a result of Blackadar and Handelman (1982) and a result of Haagerup (2014), we deduce that the set $T(A)$ is nonempty for any C^* -algebra in the class we consider.

Example 12 Let $G \curvearrowright X$ be an action of a countable discrete amenable group G on a compact metrizable space X , and let μ be a G -invariant Borel probability measure. Then μ induces a

tracial state τ_μ on the reduced crossed product $C(X) \rtimes_r G$ (see Giordano et al. (2018), Example 10.1.31). In fact, if the action is free, then every tracial state on $C(X) \rtimes_r G$ has this form (see, e.g., Giordano et al. (2018), Proposition 11.1.21).

Given a tracial state $\tau \in T(A)$, define the *rank function* $d_\tau : A_+ \rightarrow [0, \infty]$ by

$$d_\tau(a) := \lim_{n \rightarrow \infty} \tau\left(a^{\frac{1}{n}}\right)$$

(see Blackadar and Handelman (1982) for more details).

Example 13 If $A = C(X)$ and τ is the tracial state induced by a Borel probability measure μ on X , then $d_\tau(f) = \mu(f^{-1}(0, \infty))$ for any f in $C(X)_+$.

One can show that if $a \precsim b$, then $d_\tau(a) \leq d_\tau(b)$ for every $\tau \in T(A)$ (see Blackadar and Handelman (1982), Theorem II.2.2). The converse is not necessarily true, and the radius of comparison, as introduced by Toms (2006), is designed to measure the failure of this “comparison property.”

In the next definition, we use the notation $M_\infty(A) := \bigcup_{n=1}^\infty M_n(A)$, where each $M_n(A)$ is embedded into $M_{n+1}(A)$ as the upper-left corner.

Definition 12 Let A be a unital stably finite exact C^* -algebra. The *radius of comparison* $\text{rc}(A)$ of A is the infimum of the strictly positive real numbers r such that for any $a, b \in M_\infty(A)_+$, if $d_\tau(a) + r < d_\tau(b)$, for all $\tau \in T(A)$, then $a \precsim b$.

The case $\text{rc}(A) = 0$, also known as *strict comparison*, is especially important for the classification of simple nuclear C^* -algebras (see section “Classification of Simple Nuclear C^* -algebras”).

Example 14 Let X be a finite CW complex. Then by Toms (2006) and Elliott and Niu (2013), $\text{rc}(C(X))$ is essentially half of the covering dimension of X . More precisely, if $\dim X$ denotes the covering dimension of X , we have the following by Elliott and Niu (2013, Corollary 1.2):

- If $\dim X$ is odd, then

$$\mathrm{rc}(C(X)) = \max \left\{ 0, \frac{\dim X - 1}{2} - 1 \right\}.$$

- If $\dim X$ is even, then

$$\frac{\dim X}{2} - 2 \leq \mathrm{rc}(C(X)) \leq \frac{\dim X}{2} - 1.$$

The following conjecture, due to Phillips and Toms (see, e.g., the introduction of Niu (2022)), predicts a precise quantitative relationship between the mean dimension of a topological dynamical system and the radius of comparison of the associated crossed product C^* -algebra.

Conjecture 1 For a free minimal action $\alpha : G \curvearrowright X$ of a countable discrete amenable group G on a compact metrizable space X ,

$$\mathrm{rc}(C(X) \rtimes_r G) = \frac{1}{2} \mathrm{mdim}(X, \alpha, G).$$

Example 15 Let (X, T) be a Giol–Kerr minimal subshift constructed in Giol and Kerr (2010). It was proved in Hines (2015) that

$$\mathrm{rc}(C(X) \rtimes_r \mathbb{Z}) \geq \frac{1}{2} \mathrm{mdim}(X, T).$$

Example 16 Elliott and Niu (2017) proved that if $\mathrm{mdim}(X, T) = 0$, then $\mathrm{rc}(C(X) \rtimes_r \mathbb{Z}) = 0$. In fact they showed that the small boundary property, which is equivalent to mean dimension zero for $\mathbb{Z}$ -actions, implies $\mathbb{Z}$ -stability of the crossed product, which in turn implies strict comparison. See section “[Classification of Simple Nuclear \$C^*\$ -algebras](#)” for more details.

Example 17 In (Phillips 2016), Phillips showed that for a general minimal homeomorphism one always has

$$\mathrm{rc}(C(X) \rtimes_r \mathbb{Z}) \leq 1 + 36 \mathrm{mdim}(X, T).$$

In a recent breakthrough (Niu 2022), Niu established one inequality in full generality for minimal homeomorphisms.

Theorem 28 (Niu 2022, Corollary 7.10). *Let α be a minimal homeomorphism on a separable compact Hausdorff space X . Then*

$$\mathrm{rc}(C(X) \rtimes_{r, \alpha} \mathbb{Z}) \leq \frac{1}{2} \mathrm{mdim}(X, \alpha).$$

Very roughly, the proof is divided in two steps. The first one is to show that for an action α of a countable discrete amenable group G , the conclusion of Theorem 28 holds if the following two conditions are met:

- (1) $\alpha : G \curvearrowright X$ has the *uniform Rokhlin property*.
- (2) $C(X) \rtimes_{r, \alpha} G$ has *Cuntz comparison for open sets*.

Then, as the second step, it is proved that these two conditions are satisfied when $G = \mathbb{Z}$. In the preprint (Niu 2019), Niu proved that these two conditions also hold for free minimal $\mathbb{Z}^d$ -actions.

In what follows, we give the precise definitions of these two conditions. The uniform Rokhlin property, as the name suggests, is reminiscent to the classical Rokhlin lemma in ergodic theory. Recall that in the Rokhlin lemma for a single transformation T , a collection $U, T(U), \dots, T^n(U)$ of disjoint subsets is often called a *tower*. Essentially the uniform Rokhlin property says that the space can be almost covered by open disjoint Rokhlin towers, and the remainder is uniformly small for all G -invariant measures. In the section entitled “[Almost Finiteness, Comparison, and the Small Boundary Property](#)”, we will remark that the *uniform Rokhlin property* is very similar to the notion of *almost finiteness in measure*. We now recall the definition of castles given in Kerr (2020, Definition 5.7).

Definition 13 *Let G be a countable discrete group and let X be a compact metrizable G -space. A collection $\{(V_i, S_i)\}_{i \in I}$ is a *castle* if each V_i is a subset of X , each S_i is a finite subset of G , and the collection*

$$\{sV_i : s \in S_i, i \in I\} \quad (1)$$

is pairwise disjoint.

In this case, the finite sets S_i are called the shapes and the subsets sV_i are called the levels.

The following definition, due to Z. Niu in Niu (2022), was originally given in terms of orbit capacity. The statement below is equivalent thanks to Proposition 3.3 of Kerr and Szabó (2020).

Definition 14 (Niu 2022, Definition 3.1). Let G be a discrete amenable group and let X be a separable compact Hausdorff G -space. We say the action $G \curvearrowright X$ has the uniform Rokhlin property (URP) if for every finite subset $K \subseteq G$, and every $\delta > 0$, there exist:

- (1) A castle $\{(V_i, S_i)\}_{i \in I}$ such that I is finite, each level is open, and each shape is (K, δ) -invariant (i.e., $|gS_i \Delta S_i|/|S_i| < \delta$ for all $g \in K$ and all $i \in I$)
- (2) $\sup \mu(X \setminus \bigcup_{i \in I} S_i V_i) < \delta$, where the supremum is taken over all G -invariant Borel probability measures μ .

Example 18 Every free minimal homeomorphism on a Cantor set has the uniform Rokhlin property (Niu 2022, Lemma 3.5). As mentioned in Niu (2022), this essentially follows from Lindenstrauss' Rokhlin-type lemma for homeomorphisms (Lindenstrauss (2000), Corollary 3.4).

Meanwhile, Cuntz comparison for open sets basically means that $C(X) \rtimes_r G$ has strict comparison for elements in the subalgebra $C(X)$. We give the precise definition here. For each open subset E of X , choose a continuous function $\varphi : X \rightarrow [0, +\infty)$ such that $\varphi^{-1}(0, +\infty) = E$. Note that by Example 10 the Cuntz equivalence class of φ depends only on the open set E .

Definition 15 (Niu 2022, Definition 4.1). Let G be a discrete group and let X be a compact Hausdorff G -space. For $\lambda \in (0, 1]$ and $m \in \mathbb{N}$, we say the action has (λ, m) Cuntz comparison for open sets if for any open subsets E, F of X with

$$\mu(E) < \lambda \mu(F)$$

for all G -invariant Borel probability measure μ , we have

$$\varphi_E \precsim \begin{pmatrix} \varphi_F & & \\ & \ddots & \\ & & \varphi_F \end{pmatrix}_{m \times m}$$

in $M_\infty(C(X) \rtimes_r G)$.

Theorem 29 (Niu 2022, Theorem 4.7). Let $\alpha : G \curvearrowright X$ be a free minimal action of a discrete amenable group on a separable compact Hausdorff space that has (URP) and (λ, m) Cuntz comparison for open sets for some $\lambda \in (0, 1]$ and $m \in \mathbb{N}$. Then

$$\text{rc}(C(X) \rtimes_r G) \leq \frac{1}{2} \text{mdim}(X, \alpha, G).$$

Classification of Simple Nuclear C^* -algebras

This subsection contains a very brief introduction to the classification of simple nuclear C^* -algebras, or the so-called *Elliott's program*. Since the main purpose is to give more context to the preceding section on “Mean Dimension and Radius of Comparison” and the next section on “Almost finiteness, Comparison, and the Small Boundary Property”, the discussion will be rather informal.

The aim of Elliott's program is to classify the class of (unital) separable simple nuclear C^* -algebras, which includes any crossed product $C(X) \rtimes_r G$ arising from a free minimal action of a countable discrete amenable group on a compact metrizable space. The algebraic invariant is called the *Elliott invariant*, which, roughly speaking, consists of K -theory and traces. However, due to the presence of examples such as Villadsen (1998) and Rørdam (2003), and the aforementioned crossed products constructed by Giol and Kerr, we have to content ourselves with classifying “regular” C^* -algebras. One way to capture this regularity is through a noncommutative covering dimension, known as the *nuclear dimension*, introduced by Winter and Zacharias (2010). Then the regularity appears in the finiteness of the nuclear dimension.

The following classification theorem is due to many hands, including Gong et al. (2020a and 2020b); Kirchberg (1995); Phillips (2000); and Tikuisis et al. (2017) (which in turn depend on a large body of work).

Theorem 30 *Unital separable simple nuclear infinite-dimensional C^* -algebras that have finite nuclear dimension and satisfy the UCT are classified by the Elliott invariant.* (UCT stands for the *Universal Coefficient Theorem* of Rosenberg and Schochet (1987)).

In Tu (1999), the author proves that if G is an amenable group, then every crossed product $C(X) \rtimes_r G$ satisfies the UCT. It is an open question whether every nuclear C^* -algebra satisfies the UCT.

Example 19 Let θ be an irrational number in $[0, \frac{1}{2}]$ and $\alpha_\theta : \mathbb{T} \rightarrow \mathbb{T}$ be the rotation by θ on the circle. The crossed product $C(\mathbb{T}) \rtimes_{\alpha_\theta, r} \mathbb{Z}$ is known as the *irrational rotation algebra* and often denoted by $\mathcal{A}_\theta$. It can be shown that $\mathcal{A}_\theta$ has nuclear dimension one (see Winter and Zacharias (2010), Example 6.1) and belongs to the class of algebras in Theorem 30. Thanks to the work of Pimsner and Voiculescu (1980) and Rieffel (1981), $\mathcal{A}_\theta$ and $\mathcal{A}_{\theta'}$ are isomorphic if and only if $\theta = \theta'$.

Examples and results show that regularity also assumes other forms. One of them is the property of *strict comparison* (i.e., zero radius of comparison) which already appeared in the section “[Mean Dimension and Radius of Comparison](#)” (see Definition 12). Another one is known as $\mathcal{Z}$ -*stability* or *Jiang-Su stability*, which says that the C^* -algebra must remain unchanged (up to isomorphism) after tensoring with the specific C^* -algebra $\mathcal{Z}$, called the *Jiang-Su algebra* (Jiang and Su 1999). One of the important features of $\mathcal{Z}$ is that although it is infinite dimensional as a linear space, it has the same Elliott invariant as the algebra of complex numbers $\mathbb{C}$. Therefore, in some sense the Elliott invariant does not “see” the algebra $\mathcal{Z}$, whence cannot distinguish a C^* -algebra A from the tensor product $A \otimes \mathcal{Z}$.

It is natural to ask how these properties are related to each other. The following conjecture is due to Toms and Winter.

Conjecture 1 For an infinite-dimensional unital separable simple nuclear C^* -algebra, the following three conditions are equivalent:

- (i) Finite nuclear dimension
- (ii) $\mathcal{Z}$ -stability
- (iii) Strict comparison.

The implication (i) $\Rightarrow$ (ii) is proved by Winter (2012) and (ii) $\Rightarrow$ (iii) is due to Rørdam (2004). For the upward implications, many results were obtained based on the groundbreaking work of Matui and Sato (2012, 2014). In another recent breakthrough (Castillejos et al. 2021), the important notion of *uniform property Γ* was identified and the equivalence between (i) and (ii) is now fully established. Combining all the known implications we arrive at the following theorem:

Theorem 31 *Let A be a unital separable simple nuclear infinite-dimensional C^* -algebra. The following are equivalent:*

- (i) A has finite nuclear dimension.
- (ii) A is $\mathcal{Z}$ -stable.
- (iii) A has strict comparison and uniform property Γ .

In particular, the Toms–Winter conjecture holds for any unital separable simple nuclear C^ -algebra with uniform property Γ .*

At the time of writing, it is an open question whether every unital separable nuclear C^* -algebra has uniform property Γ .

Example 20 Let X be an infinite compact metrizable space, and let $T : X \rightarrow X$ be a minimal homeomorphism. Toms and Winter proved that if X has finite covering dimension, then the reduced crossed product $C(X) \rtimes_T \mathbb{Z}$ has finite nuclear dimension (and hence is $\mathcal{Z}$ -stable) (Toms and Winter 2013). Later it was shown by Elliott and

Niu that $C(X) \rtimes_{\mathbb{T}} \mathbb{Z}$ is $\mathbb{Z}$ -stable whenever (X, T) has mean dimension zero (Elliott and Niu 2017).

Almost Finiteness, Comparison, and the Small Boundary Property

A central problem at the interface of C^* -algebras and topological dynamics is to understand when a crossed product $C(X) \rtimes_{\mathbb{T}} G$ is regular, meaning that it satisfies one of the properties appearing in the Toms–Winter conjecture (Conjecture 1). Several dimension theories for topological dynamics, including Rokhlin dimension (e.g., Hirshberg et al. (2015); Szabó (2015); Szabó et al. (2019)), dynamic asymptotic dimension (Guentner et al. 2017), and tower dimension (Kerr 2020), have been introduced and they have direct applications in establishing the finiteness of nuclear dimension. Note that the dynamic asymptotic dimension was originally motivated by the Baum–Connes conjecture. In addition to applications to C^* -algebras, they are also interesting dynamical properties in their own rights. We apologize for not being able to discuss these important developments due to length, but to the interested readers we recommend the notes by Sims et al. (2020, Part III). In this subsection, we focus on the dynamical analogue of $\mathbb{Z}$ -stability and strict comparison and discuss their relationship with the small boundary property and the Toms–Winter conjecture.

We start with a dynamical version of strict comparison. The basic idea of comparing sets by their measures appeared in the work of Glasner and Weiss (1995) (and the type of subequivalence relation used there dates back to Hopf (1932)). The following definitions first appeared in talks given by Wilhelm Winter.

Definition 16 *Let G be a discrete group and X be a G -space. For two open subsets A and B of X , we say A is subequivalent to B , written as $A \prec B$, if for every closed subset C of A there are group elements $s_1, s_2, \dots, s_n$ in G and open sets $U_1, U_2, \dots, U_n$ such that $C \subseteq \bigcup_{i=1}^n U_i$, the sets $s_i U_i$ are pairwise disjoint, and $\bigcup_{i=1}^n s_i U_i \subseteq B$.*

It is clear that if $A \prec B$, then $\mu(A) \leq \mu(B)$ for every G -invariant probability measure μ .

Basically, comparison is a property that asserts the converse.

Definition 17 *Let G be a discrete group and X be a G -space. We say the action $G \curvearrowright X$ has (dynamical) comparison if $\mu(A) < \mu(B)$ for all G -invariant measures μ implies $A \prec B$ for all open sets A and B .*

Example 21 By Glasner and Weiss (1995, Lemma 2.5), every Cantor minimal system has comparison.

It's worth noting that this result essentially follows from the existence of Kakutani–Rokhlin tower decomposition for Cantor minimal systems (see the proof of Putnam (1989), Lemma 3.1), which gives us a hint to a result below that says almost finiteness implies comparison, since almost finiteness (see Definition 19 below) can be viewed as a generalization of the Kakutani–Rokhlin decomposition.

Example 22 More generally, let G be a countable discrete group of subexponential growth. A result of Downarowicz and Zhang shows that every free action of G on any zero-dimensional compact metrizable space has comparison.

Remark 21 To the authors' best knowledge, it is an open question whether every free action of a countable discrete amenable group on any compact metrizable space has comparison.

We now turn to almost finiteness, which can be viewed as a topological version of the Ornstein–Weiss tiling theorem and, at the same time, a dynamical analogue of $\mathbb{Z}$ -stability.

With the notion of castles introduced in Definition 13, we can now define almost finiteness.

Definition 18 *Let $G \curvearrowright X$ be a free action of a countable discrete group G on a compact metric space X . The action $G \curvearrowright X$ is almost finite if for every finite subset $K \subseteq G$, and every $\delta > 0$, there exist:*

- (1) *A castle $\{(V_i, S_i)\}_{i \in I}$ such that each level is open with diameter at most δ , and each shape*

is (K, δ) -invariant (i.e., $|gS_i \Delta S_i|/|S_i| < \delta$ for all $g \in K$ and all $i \in I$)

- (2) A set $S'_i \subseteq S_i$ for each $i \in I$ such that $|S'_i| < \delta|S_i|$ and

$$X \setminus \coprod_{i \in I} S_i V_i \subset \coprod_{i \in I} S'_i V_i \quad (2)$$

This is reminiscent of the celebrated Ornstein–Weiss tower decomposition (Ornstein and Weiss 1987) from ergodic theory. Essentially, almost finiteness means that except for a small part, the space can be covered by disjoint open towers whose levels have small diameter. Here the “smallness” of the remainder is captured by the notion of dynamical subequivalence, introduced in Definition 16.

Note that if an action of a group G on a space X is almost finite, then G is amenable by the (K, δ) -invariance requirement in condition (i). Therefore, in some sense almost finiteness can be viewed as a dynamical analogue of the conjunction of nuclearity and $\mathcal{Z}$ -stability.

Remark 22 The term “almost finiteness” was first introduced by Matui. In Matui (2012), Matui defined almost finiteness for étale groupoids whose unit space is compact and totally disconnected and studied the homology groups of these groupoids. It was shown in Kerr (2020) that this definition is equivalent to Definition 18 for transformation groupoids arising from groups acting on compact totally disconnected spaces.

Example 23 The Kakutani–Rokhlin tower decomposition shows that every Cantor minimal system is almost finite. Matui proved in Matui (2012) that the same holds for every free action of $\mathbb{Z}^n$ on a compact metrizable totally disconnected space.

Example 24 By Kerr and Szabó (2020, Theorem B), the results from the previous example can be extended to actions of $\mathbb{Z}^n$ on any finite-dimensional compact metrizable spaces.

The next theorem of Kerr relates almost finiteness to comparison and $\mathcal{Z}$ -stability of the associated crossed product.

Theorem 32 (Kerr 2020, Theorem 9.2 and Theorem 12.4). *Let $\alpha : G \curvearrowright X$ be a free minimal action of a countable discrete amenable group on a compact metrizable space.*

- (1) *If the action α is almost finite, then it has comparison.*
- (2) *If the action α is almost finite and G is infinite, then $C(X) \rtimes_\alpha G$ is $\mathcal{Z}$ -stable.*

Example 25 Since almost finiteness implies that the crossed product $C(X) \rtimes_\alpha G$ is $\mathcal{Z}$ -stable, the minimal actions constructed by Giol and Kerr (2010) are not almost finite.

Let us explicitly state the following corollary, which can be deduced from Theorem 32, Theorem 31, and Theorem 30. It shows that establishing almost finiteness, a purely dynamical property, directly contributes to the realm of C^* -algebras by providing more examples of classifiable algebras.

Corollary 5 *Let $\alpha : G \curvearrowright X$ be a free minimal action of a countable discrete amenable group on a compact metrizable space. If the action is almost finite, then the crossed product $C(X) \rtimes_\alpha G$ is classified by its Elliott Invariant.*

Another very natural way to quantify condition (ii) in Definition 18 is through invariant measures. This leads to the following version of almost finiteness, given in Kerr and Szabó (2020).

Definition 19 (Kerr and Szabó 2020) Definition 3.5 and Proposition 3.3) *Let G be a discrete group and let X be a compact metrizable G -space. Then, an action $G \curvearrowright X$ is almost finite in measure if it satisfies the condition (i) in Definition 18 and the following condition (ii'):*

$$\mu \left(X \setminus \coprod_{i \in I} S_i V_i \right) < \delta$$

where the supremum is taken over all G -invariant Borel probability measures μ .

If we allow, in the above definition, the levels to have arbitrary diameter then we obtain precisely the uniform Rokhlin property discussed in the section “[Mean Dimension and Radius of Comparison](#).”

Intuitively, almost finiteness implies almost finiteness in measure, and comparison allows going backward. This turns out to be true (see the proof of Kerr and Szabó (2020), Theorem 6.1), so by Theorem 32 comparison together with almost finiteness in measure implies that the crossed product is $\mathbb{Z}$ -stable. To state another characterization of almost finiteness in measure, let us recall the definition of the small boundary property introduced by Lindenstrauss and Weiss (2000).

Definition 20 *Let G be a discrete amenable group and X be a G -space. An action $G \curvearrowright X$ has the small boundary property if for every point $x \in X$ and every open neighborhood U of x , there is an open set V such that $x \in V \subseteq U$ and $\mu(\overline{V} \setminus V) = 0$ for every G -invariant probability measure μ .*

The definition of the small boundary property was first given by Lindenstrauss and Weiss using the notion of *orbit capacity* (Lindenstrauss and Weiss (2000), Definition 5.1) and was proved to be equivalent to the above definition in Proposition 3.3 of Kerr and Szabó (2020).

Remark 23 Recall that a topological space X has zero (small) inductive dimension if for every $x \in X$ and every open set U containing x , there exists an open set V such that $x \in V \subseteq U$ and V has empty boundary (see, e.g., Pears (1975), Chap. 4). Therefore, the small boundary property can be viewed as a dynamical analogue of a zero-dimensional space where the smallness of the boundary is captured by the G -invariant measures.

It was shown in Lindenstrauss and Weiss (2000) that the small boundary property implies mean dimension zero. The converse is known to be true for $\mathbb{Z}^n$ -actions (see Gutman et al. (2016);

the case of $\mathbb{Z}$ -actions was first established in Lindenstrauss (2000)). It is a major open problem whether the equivalence of these two properties still holds for actions of countable amenable groups.

The following theorem is due to Kerr and Szabó.

Theorem 33 (Kerr and Szabó (2020), Theorem 5.6). *Let $G \curvearrowright X$ be a free action of a countable discrete amenable group on a compact metrizable space. Then the action is almost finite in measure if and only if it has the small boundary property.*

As a consequence of Theorem 33 and Theorem 32, the following are equivalent for a free action of a countable amenable group on a compact metrizable space:

- (1) The action is almost finite.
- (2) The action has comparison and the small boundary property.

It might be interesting to compare this result to the equivalence between (ii) and (iii) in Theorem 31.

Returning to the connection to the classification of nuclear C^* -algebras, we recall that by Theorem 31 the Toms–Winter conjecture holds for any simple crossed product $C(X) \rtimes_r G$ that has uniform property Γ . The next result establishes a formal link between the small boundary property and uniform property Γ .

Theorem 34 (Kerr and Szabó (2020), Theorem 9.4). *Let G be an infinite countable discrete amenable group and let X be a compact metrizable G -space. If the action $G \curvearrowright X$ has the small boundary property, then the crossed product $C(X) \rtimes_r G$ has uniform property Γ .*

As a consequence, if mean dimension zero of the action turns out to imply strict comparison for the crossed product (note that this is a very special case of Conjecture 1), then every free minimal action $G \curvearrowright X$ with the small boundary property

gives rise to a $\mathbb{Z}$ -stable, and hence classifiable crossed product. Note that the same conclusion can also be reached if every such action has comparison, which is another open question (see Remark 21). Along these lines, one can ask the following question:

Question 1 Let $\alpha : G \curvearrowright X$ be a free minimal action of a countable discrete infinite amenable group on a compact metrizable space. Does $\text{mdim}(X, \alpha, G) = 0$ imply that $C(X) \rtimes_r G$ is $\mathbb{Z}$ -stable?

Note that the question has an affirmative answer for $\mathbb{Z}^n$ -actions, by a combination of Theorem 28, Gutman et al. (2016, Corollary 5.4), and Theorem 34.

Boundary Actions and C^* -Simplicity

In 1967, Dixmier asked if every simple C^* -algebra is generated by its projections. Kadison suggested that the reduced C^* -algebra $C_r^*(\mathbb{F}_2)$ of the free group on two generators should provide an example of a simple C^* -algebra with no nontrivial projections. Powers (1975) proved that $C_r^*(\mathbb{F}_2)$ is simple and has exactly one tracial state. Since then more examples of groups with these two properties have been found, and it became a major problem to characterize the groups with either property. In particular, for a discrete group G , it was open for a long time whether simplicity of $C_r^*(G)$ is equivalent to the unique trace property, that is, there is exactly one tracial state on $C_r^*(G)$.

Kalantar and Kennedy (2017) identified the Furstenberg boundary of a discrete group G , which is a purely dynamical notion, with an operator algebraic notion called the Hamana boundary. Using this identification, they characterized the simplicity of $C_r^*(G)$ in terms of the action of G on its Furstenberg boundary. Building upon this new insight, Breuillard, Kalantar, Kennedy, and Ozawa gave solutions to many long-standing questions surrounding C^* -simplicity and the unique trace property (Breuillard et al. 2017).

The Furstenberg Boundary

Let X be a compact Hausdorff space and G be a discrete group. Let $C(X)$ denote the C^* -algebra of all continuous functions on X , and $P(X)$ the set of probability measures on X , which is a compact convex subset of $C(X)^*$ (in the weak* topology). If X is a G -space, then $P(X)$ carries a natural G -space structure. Recall that a G -space X is *proximal* if for every pair x_1, x_2 in X , there is a net t_i in G such that $\lim_{i \rightarrow \infty} t_i x_1 = \lim_{i \rightarrow \infty} t_i x_2$. We say that X is *strongly proximal* if $P(X)$ is proximal as a G -space (see Glasner (1976)). A compact G -space X is called a *G -boundary* if it is minimal and strongly proximal, or equivalently if X is the unique minimal G -invariant closed subspace of $M(X)$.

By a result of Furstenberg (Glasner 1976, Theorem III.2.3), if G acts by affine transformations on a locally convex topological vector space and K is a G -invariant compact convex subset, then K contains a G -boundary. This shows that boundary actions are ubiquitous.

Theorem 35 (Furstenberg 1973). *Every discrete group G admits a unique universal G -boundary $\partial_F G$, called the Furstenberg boundary of G , in the sense that for every G -boundary X there is a surjective G -map from $\partial_F G$ onto X .*

Note that by the universal property of $\partial_F G$ and the paragraph before Theorem 35, for any compact G -space X , there exists a G -equivariant continuous map $b : \partial_F G \rightarrow P(X)$, called a *boundary map*.

Let X and Y be compact G -spaces. There is a one-to-one correspondence between continuous G -equivariant maps $a : Y \rightarrow P(X)$ and G -equivariant unital positive maps $\tilde{a} : C(X) \rightarrow C(Y)$ given by

$$\tilde{a}(f)(y) = \langle a(y), f \rangle,$$

where $\langle \cdot, \cdot \rangle$ is the canonical pairing between $P(X)$ and $C(X)$. Applying this correspondence to the boundary map $b : \partial_F G \rightarrow P(\beta G)$, where βG is the Stone–Čech compactification of G , we obtain a G -equivariant unital positive map $\tilde{b} : C(\beta G) \rightarrow C(\partial_F G)$. For a point $x \in \partial_F G$, we write $\text{ev}_x : C(\partial_F G) \rightarrow \mathbb{C}$ for the evaluation map. One can

show that the composition $\text{ev}_x \circ \tilde{b}$ restricts to an invariant mean for the stabilizer subgroup G_x (see Breuillard et al. (2017, Proposition 2.7)). It follows that for every $x \in \partial_F G$, the stabilizer subgroup G_x is amenable.

Day proved that every group has an amenable normal subgroup $R_a(G)$ that contains all the other amenable normal subgroups of G (Day (1957), Lemma 4.1). The group $R_a(G)$ is called the *amenable radical* of G . The previous paragraph, together with strong proximality and minimality of $\partial_F G$, yields the next theorem.

Recall that an action of G on X is a group homomorphism from G into $\text{Homeo}(X)$. The kernel of the action is by definition the kernel of this homomorphism.

Theorem 36 (Furman (2003); see also Breuillard et al. (2017), Proposition 2.8). *The kernel of the action $G \curvearrowright \partial_F G$ is precisely the amenable radical $R_a(G)$.*

In particular, if G is amenable, then $\partial_F G$ is a one-point space. In general, $\partial_F G$ is extremally disconnected, that is, every open subset of $\partial_F G$ has an open closure (Kalantar and Kennedy (2017), Remark 3.16; see also Breuillard et al. (2017), Proposition 2.4).

The Hamana Boundary

Definition 21 *A unital self-adjoint subspace of a unital C^* -algebra is called an operator system.*

Recall that a linear map $\varphi : A \rightarrow B$ between two C^* -algebras is *positive* if $\varphi(A^+) \subseteq B^+$, and *completely positive* (c.p.) if for every $n \in \mathbb{N}$, the matrix amplification $\varphi^{(n)} : M_n(A) \rightarrow M_n(B)$, $\varphi^{(n)}([a_{ij}]) = [\varphi(a_{ij})]$ is positive (see, e.g., Paulsen (2002)). Suppose A is a unital C^* -algebra and $S \subseteq A$ is an operator system. Then for every $n \in \mathbb{N}$, the operator system $M_n(S)$ inherits a norm and order structure from $M_n(A)$. Therefore, we can define c.p. maps between operator systems. An invertible and unital completely positive (u.c.p.) map between two operator systems whose inverse is also c.p. is called a *complete order isomorphism* (see Blackadar (2006), II.6.9.16). Moreover, a map $\varphi : S \rightarrow T$ between two operator systems is

said to be *completely isometric* if its matrix amplification $\varphi^{(n)} : M_n(S) \rightarrow M_n(T)$ is an isometry for every $n \in \mathbb{N}$.

Definition 22 (Hamana 1985). *Let S be an operator system and let G be a discrete group.*

- (1) *A G -action on S is a group homomorphism from G into the group of complete order isomorphisms on S . In this case, S is called a G -operator system.*
- (2) *A map $\varphi : S \rightarrow T$ between two G -operator systems is a G -equivariant map if*

$$\varphi(ga) = g\varphi(a)$$

for all $a \in S$ and $g \in G$.

Definition 23 (Hamana 1985). *An operator system $\mathcal{U}$ is said to be injective if for any inclusion $S \subseteq T$ of operator systems and any u.c.p. map $\varphi : S \rightarrow \mathcal{U}$, there exists a u.c.p. map $\tilde{\varphi} : T \rightarrow \mathcal{U}$ such that $\tilde{\varphi}(a) = \varphi(a)$ for all $a \in S$ (i.e., $\tilde{\varphi}$ is an extension of φ).*

Example 26

- (1) Arveson's extension theorem (see, e.g., Paulsen (2002), Chap. 7) shows that $B(H)$ is injective for any Hilbert space H .
- (2) If $\mathcal{V} \subseteq \mathcal{U}$ is an inclusion of operator systems such that $\mathcal{U}$ is injective and there is a u.c.p. projection from $\mathcal{U}$ onto $\mathcal{V}$, then $\mathcal{V}$ is necessarily injective. It follows that an operator system $\mathcal{V} \subseteq B(H)$ is injective if and only if there exists a u.c.p. projection from $B(H)$ onto $\mathcal{V}$.

Definition 24 *Let G be a discrete group. A G -operator system $\mathcal{U}$ is said to be G -injective if for any inclusion $S \subseteq T$ of G -operator systems and any G -equivariant u.c.p. map $\varphi : S \rightarrow \mathcal{U}$, there exists a G -equivariant u.c.p. map $\tilde{\varphi} : T \rightarrow \mathcal{U}$ such that $\tilde{\varphi}(a) = \varphi(a)$ for all $a \in S$ (i.e., $\tilde{\varphi}$ is an extension of φ).*

Example 27 Let $\mathcal{V}$ be an injective operator system and let G be a discrete group. The operator system $\ell^\infty(G, \mathcal{V})$ admits a G -action defined by

$$s.f(g) = f(s^{-1}g) \quad (f \in \ell^\infty(G, \mathcal{V}), s, g \in G)$$

By Hamana (1985, Lemma 2.2), $\ell^\infty(G, \mathcal{V})$ is G -injective.

Given a G -operator system $\mathcal{S}$, we are interested in finding a minimal G -injective operator system that contains $\mathcal{S}$.

Definition 25 Let G be a discrete group and $\mathcal{S}$ be a G -operator system. (Kalantar and Kennedy 2017, p. 5).

- (1) A G -extension of $\mathcal{S}$ is a pair $(\mathcal{T}, \kappa)$ of a G -operator system $\mathcal{T}$ and a G -equivariant unital complete isometry $\kappa : \mathcal{S} \rightarrow \mathcal{T}$.
- (2) A G -extension $(\mathcal{T}, \kappa)$ is said to be
 - G -injective if $\mathcal{T}$ is G -injective;
 - G -essential if for every G -equivariant u.c.p. map $\varphi : \mathcal{T} \rightarrow \mathcal{W}$, φ is a complete isometry whenever $\varphi \circ \kappa$ is

Example 28 Let $\mathcal{S} \subseteq B(H)$ be a G -operator system. Consider the map $\kappa : \mathcal{S} \rightarrow \ell^\infty(G, B(H))$ defined by

$$[\kappa(s)](g) := g^{-1}s \quad (s \in \mathcal{S}, g \in G).$$

Then the pair $(\ell^\infty(G, B(H)), \kappa)$ is a G -extension of $\mathcal{S}$, and by examples 26 and 27, this extension is G -injective.

Definition 26 Let $\mathcal{S}$ be a G -operator system. A G -extension of $\mathcal{S}$ that is both G -injective and G -essential is called a G -injective envelope of $\mathcal{S}$.

Theorem 37 (Hamana (1985), Theorem 2.5). Every G -operator system has a G -injective envelope, which is unique up to G -equivariant complete order isomorphism.

Following Hamana (1985), we write $I_G(\mathcal{S})$ for the G -injective envelope of a G -operator system $\mathcal{S}$.

We now focus on the case $\mathcal{S} = \mathbb{C}$ (equipped with the trivial action of G). Since $\ell^\infty(G) = \ell^\infty(G, \mathbb{C})$ is G -injective, the inclusion map $\mathbb{C} \rightarrow \ell^\infty(G)$ extends to a G -equivariant u.c.p. map $\varphi :$

$I_G(\mathbb{C}) \rightarrow \ell^\infty(G)$. By G -essentiality of $I_G(\mathbb{C})$, the map φ is a complete isometry, so we identify $I_G(\mathbb{C})$ with its image in $\ell^\infty(G)$. Since $I_G(\mathbb{C})$ is G -injective, the identity map on $I_G(\mathbb{C})$ extends to a (G -equivariant) u.c.p. projection

$$\psi : \ell^\infty(G) \rightarrow I_G(\mathbb{C}).$$

It follows that $I_G(\mathbb{C})$ is an injective operator system.

By Choi and Effros (1977, Theorem 3.1), the injective envelope $I_G(\mathbb{C})$ becomes a C^* -algebra when equipped with the Choi-Effros product

$$a \circ b := \psi(ab).$$

It is clear that the product is commutative, hence by the Gelfand–Naimark theorem $I_G(\mathbb{C})$ is isomorphic to the algebra of continuous functions on some compact Hausdorff space.

Definition 27 Let G be a discrete group. The Hamana boundary $\partial_H G$ is the compact Hausdorff space that satisfies $C(\partial_H G) = I_G(\mathbb{C})$.

By the Gelfand–Naimark correspondence, the action of G on $I_G(\mathbb{C})$ induces a G -action on $\partial_H G$.

Proposition 11 (Kalantar and Kennedy (2017), Proposition 3.4 and Proposition 3.7). The G -action on $\partial_H G$ is minimal and strongly proximal, that is, $\partial_H G$ is a G -boundary.

Theorem 38 (Kalantar and Kennedy (2017), Theorem 3.11). $\partial_H G = \partial_F G$.

C^* -Simplicity and Unique Trace Property

Let G be a discrete group. The left regular representation $\lambda : G \rightarrow B(\ell^2(G))$ is defined by $\lambda_g(\delta_s) = \delta_{gs}$, where δ_g is the point mass. The linear span of $\{\lambda_g : g \in G\}$ forms a $*$ -subalgebra of $B(\ell^2(G))$. Its norm closure is called the reduced group C^* -algebra of G and will be denoted by $C_r^*(G)$. We say G is C^* -simple if $C_r^*(G)$ is a simple C^* -algebra. Note that $C_r^*(G)$ has a canonical tracial state

$\tau_\lambda(a) = \langle a\delta_g, \delta_g \rangle$. We say G has the *unique trace property* if τ_λ is the only tracial state on $C_r^*(G)$.

The following groundbreaking result was first proved by Kalantar and Kennedy (2017, Theorem 6.2) and later reproved by Breuillard, Kalantar, Kennedy, and Ozawa (2017, Theorem 3.1).

Theorem 39 *Let G be a discrete group. Then the following are equivalent:*

- (1) G is C^* -simple.
- (2) G acts topologically freely on its Furstenberg boundary $\partial_F G$.
- (3) There exists a topologically free G -boundary.

Theorem 39 allows one to deduce C^* -simplicity of a group by exhibiting a topologically free boundary action.

Example 29 Let Γ be a word hyperbolic group (such as the free group $\mathbb{F}_2$). The action of Γ on its Gromov boundary $\partial\Gamma$ can be shown to be a topologically free boundary action (Ghys and de la Harpe 1988; Laca and Spielberg 1996), whence by Theorem 39 the group Γ is C^* -simple.

Although we will not discuss the proof of Theorem 39, it is worth mentioning that the crossed product $C(\partial_F G) \rtimes_r G$ played an important role in both Kalantar and Kennedy (2017) and Breuillard et al. (2017). The techniques developed in Breuillard et al. (2017) also led to the following theorem, which answers a long-standing open question.

Theorem 40 (Breuillard et al. (2017), Corollary 4.3). *A discrete group G has the unique trace property if and only if its amenable radical $R_a(G)$ is trivial.* (Haagerup (2017) gave another proof of this result.)

Combining Theorem 40, Theorem 39, and Theorem 36 yields the following corollary and hence resolves another long-standing open question. It should be noted that one can also deduce the triviality of $R_a(G)$ directly from simplicity of the C^* -algebra $C_r^*(G)$ (see de la Harpe (2007), Proposition 3).

Corollary 6 *Every discrete C^* -simple group has the unique trace property.* (The converse was later shown to be false by Le Boudec (2017).)

Theorem 39 also led to algebraic criteria for C^* -simplicity. Following Breuillard et al. (2017) we say a subgroup H of G is *normalish* if for any $n \geq 1$ and $t_1, \dots, t_n \in G$, the intersection $\cap_i t_i H t_i^{-1}$ is infinite.

Theorem 41 (Breuillard et al. (2017), Theorem 6.2). *A discrete group G is C^* -simple if it has no nontrivial finite normal subgroups and no amenable normalish subgroups.*

Using Theorem 41, Breuillard et al. (2017) not only recovered a large class of previously known C^* -simple groups but also exhibited many new ones. For example, if a countable group G admits a free probability measure-preserving action with cost strictly larger than 1 (in the sense of Gaboriau (2000)) and has no nontrivial finite normal subgroups, then G is C^* -simple (see Breuillard et al. (2017, Corollary 6.8)).

Kennedy (2020) obtained an intrinsic characterization of C^* -simplicity in terms of uniformly recurrent subgroups. The notion of uniformly recurrent subgroups is introduced by Glasner and Weiss (2015). For a discrete group G , let $\mathcal{S}(G)$ be the set of all subgroups of G . We equip $\mathcal{S}(G)$ with the Chabauty topology, which in this case coincides with the subspace topology inherited from the product space $\{0, 1\}^G$. The space $\mathcal{S}(G)$ has a G -space structure via the conjugation action $H \mapsto gHg^{-1}$. A *uniformly recurrent subgroup* is a minimal G -space $X \subseteq \mathcal{S}(G)$.

Theorem 42 (Kennedy (2020), Theorem 4.1). *A discrete group G is C^* -simple if and only if it has no nontrivial amenable uniformly recurrent subgroups.*

Using this result, one can obtain a more algebraic characterization. Following Kennedy (2020), we say a subgroup H of G is *residually normal* if there is a finite subset $F \subseteq G \setminus \{e\}$ such that the intersection $F \cap gHg^{-1}$ is nonempty for every $g \in G$.

Theorem 43 (Kennedy (2020), Theorem 5.1). *A discrete group G is C^* -simple if and only if it has no amenable residually normal subgroups.*

The results and techniques obtained in Kalantar and Kennedy (2017) and Breuillard et al. (2017) have led to many further developments and research directions, including, but not limited to, the study of ideal structure for crossed products and groupoid C^* -algebras (Borys 2019; Bryder 2017; Kalantar and Scarparo 2021; Kawabe 2017; Kennedy et al. 2021; Kennedy and Schafhauser 2019), amenability of Thompson's groups (Le Boudec and Bon 2018), and stationary C^* -dynamical systems (Hartman and Kalantar 2017).

References

- Alvin L, Ash DD, Ormes NS (2018) Bounded topological speedups. *Dyn Syst* 33(2):303–331
- Ara P, Lledó F, Perera F, eds (2011) Aspects of operator algebras and applications, volume 534 of Contemporary Mathematics. American Mathematical Society, Providence; Real Sociedad Matemática Española, Madrid. Papers from the UIMP-RSME Lluís A. Santaló Summer School in Mathematics held at the Universidad Internacional Menéndez Pelayo, Santander, July 21–25, 2008
- Arnoux P, Ornstein DS, Weiss B (1985) Cutting and stacking, interval exchanges and geometric models. *Israel J Math* 50(1–2):160–168
- Ash DD (2016) Topological speedups. ProQuest LLC, Ann Arbor, MI, Thesis (Ph.D.), University of Denver
- Bezuglyi S, Karpel O (2016) Bratteli diagrams: structure, measures, dynamics. In: Dynamics and numbers, volume 669 of Contemporary Mathematics. American Mathematical Society, Providence, pp 1–36
- Bezuglyi S, Dooley AH, Medynets K (2005) The Rokhlin lemma for homeomorphisms of a Cantor set. *Proc Am Math Soc* 133(10):2957–2964
- Blackadar B (2006) Operator algebras, volume 122 of Encyclopaedia of mathematical sciences. Springer, Berlin. Theory of C^* -algebras and von Neumann algebras, Operator Algebras and Non-commutative Geometry, III
- Blackadar B, Handelman D (1982) Dimension functions and traces on C^* -algebras. *J Funct Anal* 45(3):297–340
- Borys C (2019) The Furstenberg boundary of a groupoid. Preprint. arXiv:1904.10062
- Boyle MM (1983) Topological orbit equivalence and factor maps in symbolic dynamics. ProQuest LLC, Ann Arbor, MI, Thesis (Ph.D.), University of Washington
- Boyle M, Handelman D (1994) Entropy versus orbit equivalence for minimal homeomorphisms. *Pac J Math* 164(1):1–13
- Boyle M, Tomiyama J (1998) Bounded topological orbit equivalence and C^* -algebras. *J Math Soc Japan* 50(2): 317–329
- Bratteli O (1972) Inductive limits of finite dimensional C^* -algebras. *Trans Am Math Soc* 171:195–234
- Breuillard E, Kalantar M, Kennedy M, Ozawa N (2017) C^* -simplicity and the unique trace property for discrete groups. *Publ Math Inst Hautes Études Sci* 126:35–71
- Brown NP, Ozawa N (2008) C^* -algebras and finite-dimensional approximations. Graduate studies in mathematics, vol 88. American Mathematical Society, Providence
- Bryder RS (2017) Injective envelopes and the intersection property. Preprint. arXiv:1704.02723
- Carlsen TM, Ruiz E, Sims A, Tomforde M (2021) Reconstruction of groupoids and C^* -rigidity of dynamical systems. *Advanced Mathematics*, 390: Paper No. 107923, 55
- Castillejos J, Evington S, Tikuisis A, White S, Winter W (2021) Nuclear dimension of simple C^* -algebras. *Invent Math* 224(1):245–290
- Choi MD, Effros EG (1977) Injectivity and operator spaces. *J Funct Anal* 24(2):156–209
- Conley C, Jackson S, Marks A, Seward B, Tucker-Drob R (2020) Borel asymptotic dimension and hyperfinite equivalence relations. Preprint. arXiv:2009.06721
- Connes A (1976) Classification of injective factors. Cases II_1 , II_∞ , III_λ , $\lambda \neq 1$. *Ann Math* 104(1):73–115
- Connes A, Feldman J, Weiss B (1982) An amenable equivalence relation is generated by a single transformation. *Ergodic Theory Dynam Syst*, 1(4):431–450
- Coornaert M (2015) Topological dimension and dynamical systems. Universitext. Springer, Cham. Translated and revised from the 2005 French original
- Cortez MI (2006) $\mathbb{Z}^d$ Toeplitz arrays. *Discrete Contin Dyn Syst* 15(3):859–881
- Cortez MI, Medynets K (2016) Orbit equivalence rigidity of equicontinuous systems. *J Lond Math Soc* 94(2): 545–556
- Cortez MI, Petite S (2014) Invariant measures and orbit equivalence for generalized Toeplitz subshifts. *Groups Geom Dyn* 8(4):1007–1045
- Crainic M, Moerdijk I (2000) A homology theory for étale groupoids. *J Reine Angew Math* 521:25–46
- Cuntz J (1978) Dimension functions on simple C^* -algebras. *Math Ann* 233(2):145–153
- Dahl H (2008) AF equivalence relations associated to locally finite groups. *J Ramanujan Math Soc* 23(1): 77–95
- Day MM (1957) Amenable semigroups. *Ill J Math* 1: 509–544
- de Cornulier Y (2014) Groupes pleins-topologiques (d'après Matui, Juschenko, Monod, ...). *Astérisque*, (361):Exp. No. 1064, viii, pp 183–223
- de la Harpe P (2007) On simplicity of reduced C^* -algebras of groups. *Bull Lond Math Soc* 39(1):1–26

- Dougherty R, Jackson S, Kechris AS (1994) The structure of hyperfinite Borel equivalence relations. *Trans Am Math Soc* 341(1):193–225
- Downarowicz T (2005) Survey of odometers and Toeplitz flows. In: *Algebraic and topological dynamics*, volume 385 of *Contemporary Mathematics*. American Mathematical Society, Providence, pp 7–37
- Downarowicz T, Zhang G (2017) The comparison property of amenable groups. *arXiv:1901.01457*
- Downarowicz T, Zhang G (2019) Symbolic extensions of amenable group actions and the comparison property. *arXiv:1712.05129*
- Durand F (2010) Combinatorics on Bratteli diagrams and dynamical systems. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.* Cambridge University Press, Cambridge, pp 324–372
- Durand F, Perrin D (2022) Dimension groups and dynamical systems—substitutions, Bratteli diagrams and Cantor systems, volume 196 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge
- Dye HA (1959) On groups of measure preserving transformations. I. *Am J Math* 81:119–159
- Dye HA (1963) On groups of measure preserving transformations. II. *Am J Math* 85:551–576
- Effros EG (1981) Dimensions and C^* -algebras, volume 46 of *CBMS Regional conference series in Mathematics*. Conference Board of the Mathematical Sciences, Washington, DC
- Effros EG, Handelmann DE, Shen CL (1980) Dimension groups and their affine representations. *Am J Math* 102(2):385–407
- Elliott GA (1976) On the classification of inductive limits of sequences of semisimple finite-dimensional algebras. *J Algebra* 38(1):29–44
- Elliott GA, Niu Z (2013) On the radius of comparison of a commutative C^* -algebra. *Can Math Bull* 56(4):737–744
- Elliott GA, Niu Z (2017) The C^* -algebra of a minimal homeomorphism of zero mean dimension. *Duke Math J* 166(18):3569–3594
- Ellis R (1960) Universal minimal sets. *Proc Am Math Soc* 11:540–543
- Feldman J, Moore CC (1977) Ergodic equivalence relations, cohomology, and von Neumann algebras. I. *Trans Am Math Soc* 234(2):289–324
- Forrest AH, Hunton J (1999) The cohomology and K-theory of commuting homeomorphisms of the Cantor set. *Ergodic Theory Dynam Syst* 19(3):611–625
- Furman A (1999) Orbit equivalence rigidity. *Ann Math* 150(3):1083–1108
- Furman A (2003) On minimal strongly proximal actions of locally compact groups. *Israel J Math* 136:173–187
- Furstenberg H (1973) Boundary theory and stochastic processes on homogeneous spaces. In *Harmonic analysis on homogeneous spaces (Proc. Sympos. Pure Math., Vol. XXVI, Williams Coll., Williamstown, MA., 1972)*, pp 193–229
- Gaboriau D (2000) Coût des relations d'équivalence et des groupes. *Invent Math* 139(1):41–98
- Gähler F, Hunton J, Kellendonk J (2013) Integral cohomology of rational projection method patterns. *Algebr Geom Topol* 13(3):1661–1708
- Gao S, Jackson S (2015) Countable abelian group actions and hyperfinite equivalence relations. *Invent Math* 201(1):309–383
- Ghys É, de la Harpe P, eds (1988) *Sur les groupes hyperboliques d'après Mikhael Gromov*, volume 83 of *Progress in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1990. *Papers from the Swiss Seminar on Hyperbolic Groups held in Bern*
- Giol J, Kerr D (2010) Subshifts and perforation. *J Reine Angew Math* 639:107–119
- Giordano T, de la Harpe P (1997) Moyennabilité des groupes dénombrables et actions sur les espaces de Cantor. *C R Acad Sci* 324(11):1255–1258
- Giordano T, Putnam IF, Skau CF (1995) Topological orbit equivalence and C^* -crossed products. *J Reine Angew Math* 469:51–111
- Giordano T, Putnam IF, Skau CF (1999) Full groups of Cantor minimal systems. *Israel J Math* 111:285–320
- Giordano T, Putnam I, Skau C (2004) Affable equivalence relations and orbit structure of Cantor dynamical systems. *Ergodic Theory Dynam Syst* 24(2):441–475
- Giordano T, Matui H, Putnam IF, Skau CF (2008) Orbit equivalence for Cantor minimal $\mathbb{Z}^2$ -systems. *J Am Math Soc* 21(3):863–892
- Giordano T, Matui H, Putnam IF, Skau CF (2010) Orbit equivalence for Cantor minimal $\mathbb{Z}^d$ -systems. *Invent Math* 179(1):119–158
- Giordano T, Kerr D, Phillips NC, Toms A (2018) Crossed products of C^* -algebras, topological dynamics, and classification. *Advanced Courses in Mathematics*. CRM Barcelona. Birkhäuser/Springer, Cham. Lecture notes based on the course held at the Centre de Recerca Matemàtica (CRM) Barcelona, June 14–23, 2011 Edited by Francesc Perera
- Giordano T, Putnam IF, Skau CF (2019) $\mathbb{Z}^d$ -odometers and cohomology. *Groups Geom Dyn* 13(3):909–938
- Glasner S (1976) Proximal flows. *Lecture notes in Mathematics*, vol 517. Springer, Berlin/New York
- Glasner E (2003) *Ergodic theory via joinings*, volume 101 of *Mathematical surveys and monographs*. American Mathematical Society, Providence
- Glasner E, Weiss B (1995) Weak orbit equivalence of Cantor minimal systems. *Int J Math* 6(4):559–579
- Glasner E, Weiss B (2015) Uniformly recurrent subgroups. In: *Recent trends in ergodic theory and dynamical systems*, volume 631 of *Contemp. Math.* American Mathematical Society, Providence, pp 63–75
- Gong G, Lin H, Niu Z (2020a) A classification of finite simple amenable $\mathbb{Z}$ -stable C^* -algebras, I: C^* -algebras with generalized tracial rank one. *C R Math Acad Sci Soc R Can* 42(3):63–450
- Gong G, Lin H, Niu Z (2020b) A classification of finite simple amenable $\mathbb{Z}$ -stable C^* -algebras, II: C^* -algebras with rational generalized tracial rank one. *C R Math Acad Sci Soc R Can* 42(4):451–539
- Guentner E, Willett R, Guoliang Y (2017) Dynamic asymptotic dimension: relation to dynamics, topology,

- coarse geometry, and C^* -algebras. *Math Ann* 367(1–2):785–829
- Gutman Y, Lindenstrauss E, Tsukamoto M (2016) Mean dimension of $\mathbb{Z}^k$ -actions. *Geom Funct Anal* 26(3):778–817
- Haagerup U (1987) Connes' bicentralizer problem and uniqueness of the injective factor of type III_1 . *Acta Math* 158(1–2):95–148
- Haagerup U (2014) Quasitraces on exact C^* -algebras are traces. *C R Math Acad Sci Soc R Can* 36(2–3):67–92
- Haagerup U (2017) A new look at C^* -simplicity and the unique trace property of a group. In *Operator algebras and applications—the Abel Symposium 2015*, volume 12 of *Abel Symp*. Springer, Cham, pp 167–176
- Hamana M (1985) Injective envelopes of C^* -dynamical systems. *Tohoku Math J* 37(4):463–487
- Hartman Y, Kalantar M (2017) Stationary C^* -dynamical systems. To appear in *J. Eur. Math. Soc.* (includes an appendix by Uri Bader, Yair Hartman, and Mehrdad Kalantar). arXiv:2107.03980
- Herman RH, Putnam IF, Skau CF (1992) Ordered Bratteli diagrams, dimension groups and topological dynamics. *Int J Math* 3(6):827–864
- Hines T (2015) The radius of comparison and mean dimension. ProQuest LLC, Ann Arbor, MI, Thesis (Ph.D.)—Purdue University
- Hirshberg I, Winter W, Zacharias J (2015) Rokhlin dimension and C^* -dynamics. *Commun Math Phys* 335(2):637–670
- Hjorth G, Molberg M (2006) Free continuous actions on zero-dimensional spaces. *Topology Appl* 153(7):1116–1131
- Hopf E (1932) Theory of measure and invariant integrals. *Trans Am Math Soc* 34(2):373–393
- Høynes S-M (2016) Toeplitz flows and their ordered K-theory. *Ergodic Theory Dynam Syst* 36(6):1892–1921
- Hunton J (2015) Spaces of projection method patterns and their cohomology. In: *Mathematics of aperiodic order*, volume 309 of *Progr. Math.* Birkhäuser/Springer, Basel, pp 105–135
- Ioana A (2011) W^* -superrigidity for Bernoulli actions of property (T) groups. *J Am Math Soc* 24(4):1175–1226
- Ioana A (2013) Classification and rigidity for von Neumann algebras. In: *European Congress of Mathematics*. European Mathematical Society, Zürich, pp 601–625
- Ioana A (2018) Rigidity for von Neumann algebras. In: *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018*. Vol. III. Invited lectures. World Sci. Publ., Hackensack, pp 1639–1672
- Jackson S, Kechris AS, Louveau A (2002) Countable Borel equivalence relations. *J Math Log* 2(1):1–80
- Jiang X, Su H (1999) On a simple unital projectionless C^* -algebra. *Am J Math* 121(2):359–413
- Johansen Ø (1998) Ordered K-theory and Bratteli diagrams: implications for Cantor minimal systems. Ph. D. thesis, NTNU
- Johnson ASA, McClendon DM (2022) Topological speedups of $\mathbb{Z}^d$ -actions. *Dyn Syst* 37(2):222–261
- Kalantar M, Kennedy M (2017) Boundaries of reduced C^* -algebras of discrete groups. *J Reine Angew Math* 727:247–267
- Kalantar M, Scarparo E (2021) Boundary maps and covariant representations. Preprint. arXiv:2106.06382
- Katzlinger L (2019) Topological full groups. Preprint. arXiv:1907.07424
- Kawabe T (2017) Uniformly recurrent subgroups and the ideal structure of reduced crossed products. Preprint. arXiv:1701.03413
- Kennedy M (2020) An intrinsic characterization of C^* -simplicity. *Ann Sci Éc Norm Supér* 53(5):1105–1119
- Kennedy M, Schafhauser C (2019) Noncommutative boundaries and the ideal structure of reduced crossed products. *Duke Math J* 168(17):3215–3260
- Kennedy M, Kim S-J, Li X, Raum S, Ursu D (2021) The ideal intersection property for essential groupoid C^* -algebras. Preprint. arXiv:2107.03980
- Kerr D (2020) Dimension, comparison, and almost finiteness. *J Eur Math Soc (JEMS)* 22(11):3697–3745
- Kerr D, Szabó G (2020) Almost finiteness and the small boundary property. *Commun Math Phys* 374(1):1–31
- Kirchberg E (1995) Exact C^* -algebras, tensor products, and the classification of purely infinite algebras. In *Proceedings of the International Congress of Mathematicians*, vol. 1, 2 (Zürich, 1994), pp 943–954. Birkhäuser, Basel
- Krieger W (1976) On ergodic flows and the isomorphism of factors. *Math Ann* 223(1):19–70
- Krieger W (1979/80) On a dimension for a class of homeomorphism groups. *Math Ann*, 252(2):87–95
- Kuratowski K (1968) *Topology*. Vol. II. Academic Press, New York-London; Państwowe Wydawnictwo Naukowe [Polish Scientific Publishers], Warsaw. New edition, revised and augmented, Translated from the French by A. Kirkor
- Laca M, Spielberg J (1996) Purely infinite C^* -algebras from boundary actions of discrete groups. *J Reine Angew Math* 480:125–139
- Le Boudec A (2017) C^* -simplicity and the amenable radical. *Invent Math* 209(1):159–174
- Le Boudec A, Bon NM (2018) Subgroup dynamics and C^* -simplicity of groups of homeomorphisms. *Ann Sci Éc Norm Supér* 51(3):557–602
- Li H (2013) Sofic mean dimension. *Adv Math* 244:570–604
- Li X (2018a) Continuous orbit equivalence rigidity. *Ergodic Theory Dynam Syst* 38(4):1543–1563
- Li X (2018b) Dynamic characterizations of quasi-isometry and applications to cohomology. *Algebr Geom Topol* 18(6):3477–3535
- Lindenstrauss E (2000) Mean dimension, small entropy factors and an embedding theorem. *Inst Hautes Études Sci Publ Math* (89):227–262. 1999
- Lindenstrauss E, Weiss B (2000) Mean topological dimension. *Israel J Math* 115:1–24
- Matui H (2008a) An absorption theorem for minimal AF equivalence relations on Cantor sets. *J Math Soc Japan* 60(4):1171–1185

- Matui H (2008b) Torsion in coinvariants of certain Cantor minimal $\mathbb{Z}^2$ -systems. *Trans Am Math Soc* 360(9): 4913–4928
- Matui H (2012) Homology and topological full groups of étale groupoids on totally disconnected spaces. *Proc Lond Math Soc* 104(1):27–56
- Matui H (2015) Topological full groups of one-sided shifts of finite type. *J. Reine Angew. Math.* 705:35–84
- Matui H (2016) Étale groupoids arising from products of shifts of finite type. *Adv Math* 303:502–548
- Matui H (2017) Topological full groups of étale groupoids. In: *Operator algebras and applications—the Abel Symposium 2015*, volume 12 of *Abel Symposium*. Springer, Cham, pp 203–230
- Matui H, Sato Y (2012) Strict comparison and $\mathcal{Z}$ -absorption of nuclear C^* -algebras. *Acta Math.* 209(1):179–196
- Matui H, Sato Y (2014) Decomposition rank of UHF-absorbing C^* -algebras. *Duke Math J* 163(14):2687–2708
- Medynets K (2006) Cantor aperiodic systems and Bratteli diagrams. *C R Math Acad Sci Paris* 342(1):43–46
- Medynets K, Sauer R, Thom A (2017) Cantor systems and quasi-isometry of groups. *Bull Lond Math Soc* 49(4): 709–724
- Monod N, Shalom Y (2006) Orbit equivalence rigidity and bounded cohomology. *Ann Math* 164(3):825–878
- Murray FJ, von Neumann J (1936) On rings of operators. *Ann Math* 37(1):116–229
- Murray FJ, von Neumann J (1943) On rings of operators. IV. *Ann Math* 44:716–808
- Niu Z (2019) Comparison radius and mean topological dimension: $\mathbb{Z}^d$ -actions. Preprint. arXiv:1906.09172
- Niu Z (2022) Comparison radius and mean topological dimension: Rokhlin property, comparison of open sets, and subhomogeneous C^* -algebras. *J Anal Math* 146(2):595–672
- Ormes NS (1997) Strong orbit realization for minimal homeomorphisms. *J Anal Math* 71:103–133
- Ornstein DS, Weiss B (1980) Ergodic theory of amenable group actions. I. The Rohlin lemma. *Bull Amer Math Soc (NS)* 2(1):161–164
- Ornstein DS, Weiss B (1987) Entropy and isomorphism theorems for actions of amenable groups. *J Analyse Math* 48:1–141
- Paterson ALT (1999) Groupoids, inverse semigroups, and their operator algebras, volume 170 of *Progress in Mathematics*. Birkhäuser Boston, Inc, Boston, MA
- Paulsen V (2002) Completely bounded maps and operator algebras, volume 78 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge
- Pears AR (1975) *Dimension theory of general spaces*. Cambridge University Press, Cambridge, UK
- Phillips NC (2000) A classification theorem for nuclear purely infinite simple C^* -algebras. *Doc Math* 5:49–114
- Phillips NC (2005) Crossed products of the Cantor set by free minimal actions of $\mathbb{Z}^d$. *Commun Math Phys* 256(1):1–42
- Phillips NC (2016) The C^* -algebra of a minimal homeomorphism with finite mean dimension has finite radius of comparison. Preprint. arXiv:1605.07976
- Pimsner M, Voiculescu D (1980) Imbedding the irrational rotation C^* -algebra into an AF-algebra. *J Operator Theory* 4(2):201–210
- Poon YT (1989) A K -theoretic invariant for dynamical systems. *Trans Am Math Soc* 311(2):515–533
- Popa S (2007) Cocycle and orbit equivalence superrigidity for malleable actions of w -rigid groups. *Invent Math* 170(2):243–295
- Powers RT (1975) Simplicity of the C^* -algebra associated with the free group on two generators. *Duke Math. J.* 42:151–156
- Putnam IF (1989) The C^* -algebras associated with minimal homeomorphisms of the Cantor set. *Pac J Math* 136(2):329–353
- Putnam IF (2010) Orbit equivalence of Cantor minimal systems: a survey and a new proof. *Expo Math* 28(2): 101–131
- Putnam IF (2018) *Cantor minimal systems*, volume 70 of *University lecture series*. American Mathematical Society, Providence
- Putnam I, Schmidt K, Skau C (1986) C^* -algebras associated with Denjoy homeomorphisms of the circle. *J Oper Theory* 16(1):99–126
- Renault J (1980) A groupoid approach to C^* -algebras, volume 793 of *Lecture notes in Mathematics*. Springer, Berlin
- Renault J (2003) AF equivalence relations and their cocycles. In: *Operator algebras and mathematical physics (Constantza, 2001)*. Theta, Bucharest, pp 365–377
- Renault J (2008) Cartan subalgebras in C^* -algebras. *Irish Math Soc Bull* 61:29–63
- Rieffel MA (1981) C^* -algebras associated with irrational rotations. *Pac J Math* 93(2):415–429
- Rørdam M (2003) A simple C^* -algebra with a finite and an infinite projection. *Acta Math* 191(1):109–142
- Rørdam M (2004) The stable and the real rank of $\mathcal{Z}$ -absorbing C^* -algebras. *Int J Math* 15(10):1065–1084
- Rosenberg J, Schochet C (1987) The Künneth theorem and the universal coefficient theorem for Kasparov’s generalized K -functor. *Duke Math J* 55(2):431–474
- Sims A, Szabó G, Williams D (2020) *Operator algebras and dynamics: groupoids, crossed products, and Rokhlin dimension*. Advanced courses in Mathematics. CRM Barcelona. Birkhäuser/Springer, Cham, © 2020. Lecture notes from the Advanced Course held at Centre de Recerca Matemàtica (CRM) Barcelona, March 13–17, 2017, Edited by Francesc Perera.
- Singer IM (1955) Automorphisms of finite factors. *Am J Math* 77:117–133
- Skau C (2000) Ordered K -theory and minimal symbolic dynamical systems. *Colloq Math* 84/85:203–227. Dedicated to the memory of Anzelm Iwanik
- Sugisaki F (1998) The relationship between entropy and strong orbit equivalence for the minimal homeomorphisms. II. *Tokyo J Math* 21(2):311–351

- Sugisaki F (2003) The relationship between entropy and strong orbit equivalence for the minimal homeomorphisms. I. *Internat J Math* 14(7):735–772
- Szabó G (2015) The Rokhlin dimension of topological $\mathbb{Z}^m$ -actions. *Proc Lond Math Soc* 110(3):673–694
- Szabó G, Wu J, Zacharias J (2019) Rokhlin dimension for actions of residually finite groups. *Ergodic Theory Dynam Syst* 39(8):2248–2304
- Tikuisis A, White S, Winter W (2017) Quasidiagonality of nuclear C^* -algebras. *Ann Math* 185(1):229–284
- Tomiyama J (1996) Topological full groups and structure of normalizers in transformation group C^* -algebras. *Pacific J Undergrad Math* 173(2):571–583
- Toms AS (2006) Flat dimension growth for C^* -algebras. *J Funct Anal* 238(2):678–708
- Toms AS, Winter W (2013) Minimal dynamics and K-theoretic rigidity: Elliott’s conjecture. *Geom Funct Anal* 23(1):467–481
- Tu J-L (1999) La conjecture de Baum-Connes pour les feuilletages moyennables. *K-Theory* 17(3):215–264
- Vershik AM (1981) Uniform algebraic approximation of shift and multiplication operators. *Dokl Akad Nauk SSSR* 259(3):526–529
- Vershik AM (1982) A theorem on Markov periodic approximation in ergodic theory. *Zap Nauchn Sem Leningrad Otdel Mat Inst Steklov (LOMI)* 115:72–82. 306. Boundary value problems of mathematical physics and related questions in the theory of functions, 14
- Villadsen J (1998) Simple C^* -algebras with perforation. *J Funct Anal* 154(1):110–116
- von Neumann J (1932) Proof of the quasi-ergodic hypothesis. *Proc Natl Acad Sci* 18(1):70–82
- Winter W (2012) Nuclear dimension and $\mathbb{Z}$ -stability of pure C^* -algebras. *Invent Math* 187(2):259–342
- Winter W, Zacharias J (2010) The nuclear dimension of C^* -algebras. *Adv Math* 224(2):461–498

The Complexity and the Structure and Classification of Dynamical Systems

Matthew Foreman
University of California at Irvine, Irvine, CA,
USA

Article Outline

Glossary

Introduction: What Is a Dynamical System? What Is Structure? What Is a Classification?
Examples of Structure and Classification Results in Dynamical Systems
Descriptive Complexity
Complexity in Structure Theory
Complexity in Classification Theory
Standard Mathematical Objects in Each Region
Placing Dynamical Systems in Each Region
Bibliography

Glossary

Analytic set A subset A of a Polish space X is *analytic* if there is a Polish space Y and a Borel set $B \subseteq X \times Y$ such that $A = \{x: \text{for some } y \in Y, (x, y) \in B\}$. Equivalently A is the projection of B to the X -axis.

Anosov diffeomorphism Definition 13.

Benchmarking An informal term describing the location of a set or an equivalence relation in terms of reducibility to other sets or equivalence relations.

Bernoulli shift Definition 7.

Borel hierarchy Definition 70

Co-analytic set A subset C of a Polish space X is *co-analytic* if there is a Polish space Y and a Borel set $B \subseteq X \times Y$ such that $C = \{x \text{ for all } y \in Y, (x, y) \notin B\}$. Equivalently, the complement of C is analytic.

Descriptive complexity This term is used for the placement of a structure or a classification problem among the benchmarks of reducibility.

Distal *Topologically distal* is definition 20, *measure distal* is definition 27.

$=^+$ equivalence relation Introduced and discussed in section “ $=^+$ ” and the Friedman-Stanley jump operator.”

E_0 equivalence relation Definition 32.

Factor map: measurable Let X, Y be standard measure spaces and $T: X \rightarrow X, S: Y \rightarrow Y$ be measure-preserving transformations. Then S is a *factor* of T if there is a (not necessarily invertible) measure-preserving $\pi: X \rightarrow Y$ such that $\pi \circ T = S \circ \pi$ almost everywhere. The map π is the *factor map*.

Factor map: topological Let X, Y be topological spaces and $T: X \rightarrow X, S: Y \rightarrow Y$ be homeomorphisms. Then S is a *factor* of T if there is a continuous surjective map $\pi: X \rightarrow Y$ such that $\pi \circ T = S \circ \pi$. The map π is the *factor map*.

Ill-founded, Well-founded If $T \subseteq X^{<\mathbb{N}}$ is a tree, then T is *well-founded* if and only if there is no function $f: \mathbb{N} \rightarrow X$ such that for all $n, f \upharpoonright \{0, 1, 2, \dots, n\} \in T$. Equivalently, T has no infinite paths. If T is not well-founded, then T is *ill-founded*.

$\mathcal{K}$ -automorphisms Definition 10.

Kakutani equivalence Definition 60.

Minimal homeomorphism A homeomorphism $h: X \rightarrow X$ is *minimal* if for every $x \in X$, $\{h^n(x): n \in \mathbb{N}\}$ is dense.

Measure conjugacy Let S, T be measure-preserving transformations defined on standard measure spaces $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$. Then S, T are *measure conjugate* if there is an invertible measure-preserving transformations $\phi: X \rightarrow Y$ such that $\phi \circ T = S \circ \phi$ almost everywhere.

Morse-Smale diffeomorphism Definition 14.

Π_1^1 -norms Definition 75.

$\phi^*(\mu)$ Let X, Y be Polish spaces, μ a measure on X and $\phi: X \rightarrow Y$ be a measurable map. Then ϕ

and μ induce a measure $\phi^*(\mu)$ on the Borel subsets of Y by setting $\phi^*(\mu)(B) = \mu(\phi^{-1}(B))$.

Polish space A Polish space is a topological space (X, τ) such that there is some complete separable metric d on X inducing the topology τ . The topology τ is called a *Polish Topology*.

Polish group action A *Polish group* is a topological group with a Polish topology. A Polish group action is a Polish group acting G acting jointly continuously on a Polish space X . See section “Polish Group Actions.”

Reduction If A and B are subsets of Polish spaces X and Y , then a one dimensional *reduction* of A to B is a function $f: X \rightarrow Y$ such that for all $x, x \in A$ if and only if $f(x) \in B$. If $A \subseteq X \times X$ and $B \subseteq Y \times Y$, then a two dimensional reduction is a function $f: X \rightarrow Y$ such that for all $(x_1, x_2) \in X \times X, (x_1, x_2) \in A$ if and only if $(f(x_1), f(x_2)) \in B$. The reduction is *continuous* or *Borel* if the function is continuous or Borel. One- and two-dimensional Borel reductions are notated as $\leq_{\mathcal{B}}^1$ and $\leq_{\mathcal{B}}^2$. (section “What Is a Reduction?”).

Rotation number Definition 2

Separable and complete measure spaces A measure space $(X, \mathcal{B}, \mu)$ is *separable* if there is a countable subset $\mathcal{A} \subseteq \mathcal{B}$ such that for all $S \in \mathcal{B}$ and $\epsilon > 0$, there is an $A \in \mathcal{A}$ such that $\mu(A \Delta S) < \epsilon$. A measure space is *complete* if whenever $A \in \mathcal{B}$ has $\mu(A) = 0$ and $B \subseteq A$, then $B \in \mathcal{B}$.

Smooth equivalence relation Definition 31.

Smooth manifold A manifold with a C^k structure for some $k \geq 1$.

Smooth transformation A C^k -map $f: M \rightarrow M$ where M has a C^k -structure for some $k \geq 1$.

Standard measure space A measure space $(X, \mathcal{B}, \mu)$ is *standard* if it is separable and complete.

The group S_∞ The group S_∞ is the group of permutations of the natural numbers (section “ S_∞ -Actions”).

Topological conjugacy Let X, Y be topological spaces and $f: X \rightarrow X, g: Y \rightarrow Y$. Then f and g are topologically conjugate if there is a homeomorphism $\phi: X \rightarrow Y$ such that $\phi \circ f = g \circ \phi$.

Trees A (downward branching) tree is a partial ordering $(T, \leq_T)$ such that for all $t \in T$, there is a

well-ordering $\leq^*$ of $\{s \in T: s \geq t\}$ such for $s_1, s_2 \geq t, s_2 \leq^* s_1$ if and only if $s_1 \leq^* s_2$. A *path* through a tree is a one-to-one function p from an ordinal α whose range is $\leq_T$ -upwards closed and if $\beta < \gamma < \alpha, p(\gamma) \leq_T p(\beta)$. A *branch* through T is a maximal path through T . A tree is *well-founded* if and only if it has no infinite paths.

Trees of finite sequences Fix a set X . Let $X^{<\mathbb{N}}$ be the collection of finite sequences of elements of X . Order $X^{<\mathbb{N}}$ by setting $\sigma \leq \tau$ if and only if σ is an initial segment of τ . A set $T \subseteq X^{<\mathbb{N}}$ is a *tree of finite sequences of elements of X* if and only if whenever $\sigma \in T$ and τ is an initial segment of σ then $\tau \in T$. Trees of finite sequences can be viewed as elements of $\{0, 1\}^{X^{<\mathbb{N}}}$. Putting the discrete topology on $\{0, 1\}$ and the product topology on $\{0, 1\}^{X^{<\mathbb{N}}}$, the space of trees is a compact topological space. If X is countable then the space of trees on X is homeomorphic to the Cantor set. Trees are discussed in section “Reducing Ill-Founded Trees.”

Topologically transitive If X is a metric space, a homeomorphism $h: X \rightarrow X$ is topologically transitive if for some $x \in X, \{h^n(x) : n \in \mathbb{N}\}$ is dense.

Turbulent Turbulence is a property of a continuous action of a Polish group on a Polish space. It gives a method for showing that an equivalence relation is not the result of an S_∞ -action. It is discussed in section “Turbulence” where the notion is formally defined.

Well-founded, Ill-founded If $T \subseteq X^{<\mathbb{N}}$ is a tree, then T is *well-founded* if and only if there is no function $f: \mathbb{N} \rightarrow X$ such that for all $n, f \upharpoonright \{0, 1, 2, \dots, n\} \in T$. Equivalently, if T has no infinite paths, T is not well-founded, then T is *ill-founded*.

Introduction: What Is a Dynamical System? What Is Structure? What Is a Classification?

This article is a survey of the *complexity* of structure, anti-structure, classification, and anti-classification results in dynamical systems. The entry focusses primarily on ergodic theory, with excursions into topological dynamical systems.

Part of the intention is to suggest methods and problems in related areas. It is not meant to be a comprehensive study in any sense and makes no attempt to place results in a historical context. Rather it is intended to suggest classification and anti-classification results that are likely to be common in many areas of dynamical systems. Hence, the focus is on giving particular examples where descriptive set theory techniques have been successful, and mention related work in passing to suggest that the examples given here are not isolated (Apologies for leaving out important results, this reflects my ignorance and space limitations.).

The target audience for this entry is twofold. In no particular order, one audience consists of researchers whose primary interest is in some area of dynamics who may not be aware of, or who want to learn more about the methods for studying the complexity of structure and classification offered by descriptive set theory. For these people, the attempt is to give familiar examples from dynamical systems where the complexity is understood relative to known benchmarks. Logical technicalities are avoided as much as possible, and the background necessary is given in the appendix. This is *naïve* descriptive set theory – with the rough meaning that it omits the technicalities of quantifiers. A self-contained source on basic descriptive set theory, requiring very minimal background is Foreman (2010). There are many excellent, more advanced sources for this material (Kechris 1995; Marker 2002; Moschovakis 2009). The author makes no attempt at historical attributions of the theorems in descriptive set theory proper, these are well-covered in Kechris (1995); Moschovakis (2009).

The second audience are those descriptive set theorists who are likely to be very familiar with the language of *reductions* and *quantifier complexity*. For these people, the attempt is to give a very high-level overview of some of the known results specifically in dynamical systems. The choices of examples reflect the authors taste and background, but include extremely famous results and the problems left open by them. There is a list of open problems at the end of the entry and more will appear in a future arXiv submission with several co-authors, particular F. G. Ramos.

The upshot is that in any given paragraph in the entry, one of the audiences will be questioning the point of what is being presented. I ask the reader's patience in understanding the dual mission.

What Is the Difference Between a Structure Theory and a Classification? The setting for all of these results is a Polish space X , and we study classes $C \subseteq X$. In some cases $C = X$, in which case the structure question is less cogent.

The distinction between structure and classification is a bit vague. Very roughly, a structure theorem for C gives criteria for membership of elements that belong to X to be in C . The structure theorem gives more information than the definition of C .

Often this additional information simply gives a very explicit test for belonging to C that may use ideas slightly different than the definition. Another form of a structure theorem is a method for building an element of C . The form of this type of classification is:

$$x \in C$$

if and only if

x can be built using a construction with properties 1)-n)

where properties 1-n) are concrete and explicit.

While a structure theory involves one object at a time and answers the question whether that object is in $C \subseteq X$, a classification theory involves two objects known to be in C and asks whether they are equivalent with respect to an equivalence relation in question. It is usually concerned with assigning invariants to the equivalence classes.

Examples of common equivalence relations can include being in the same orbit by a group action, or being isomorphic, or have some other property in common.

This entry is concerned with two cogent questions. The first is

Can the structure or classification theory be done with inherently countable information (is it Borel)?

The second question is:

Where does its complexity sit relative to existing benchmarks? For example: Are there numerical invariants? Can one assign countable structures to elements of the class so that isomorphism is the invariant?

Many classical structure theorems were proved in the 1960s and 1970s. They fit well into the rubric suggested by this entry.

What is anti-classification? Reductions are tools that allow the establishment of lower bounds on the complexity of equivalence relations. These lower bounds are often the established benchmarks discussed in this entry. The complexity of some of some benchmarks is extremely high—often they are not even Borel.

Why is this important? An underlying thesis of this entry is that a general solution to a question about a Polish space that is not Borel cannot be solved with inherently countable information (This is not original to this paper). If an equivalence relation is not Borel a general question whether $x_1 \sim x_2$ requires *some* uncountable resource — usually an application of the uncountable Axiom of Choice. Thus, a general theme of this entry consists of determining when a question can be answered using inherently countable information: it focusses on the Borel/non-Borel distinction.

This language can be confusing: for example, if an equivalence relation E can be “classified by countable structures,” it sounds like it is “classifiable.” However even for countable structures, the isomorphism relation may not be Borel. An example of this is the equivalence relation of *isomorphism* for *countable groups* (See section “ S_∞ -Actions”).

Dynamical Systems Broadly speaking a *dynamical system* is any group action $\phi : G \times X \rightarrow X$. However, to be interesting, the group actions are taken to preserve some structure on X . Commonly studied types of structure include topological, measure theoretic, smooth, and complex.

There is a deep and well-developed theory for general groups, both amenable and non-amenable, discrete, topological, and carrying a differentiable or complex structure. Accordingly, the actions can be discrete, continuous, or smooth actions.

Typical groups include $\mathbb{Z}^n$, $\mathbb{R}^n$, more general Lie groups, free groups $\mathbb{F}_n$, and groups arising as automorphism groups of natural structures. However, much of the theory is modeled on initial

successes with the groups $\mathbb{Z}$ and $\mathbb{R}$, so that will be the focus in this entry. In general the theory gets progressively more difficult as the groups move from $\mathbb{Z}$ to $\mathbb{Z}^d$ to general amenable groups, to free groups, and then to general non-amenable groups. Since part of the story being told here is that classifications can be “impossible,” showing this in the simplest, most concrete situation illustrates the point most dramatically.

The roots of the theory of much of dynamical systems can be traced to the study of vector fields on smooth manifolds, naturally linked to smooth $\mathbb{R}$ -actions. (See Smale 1963). As argued by Smale in (1967) these actions have smooth cross sections that give significant information about the corresponding solutions to ordinary differential equations. The smooth cross-sections are $\mathbb{Z}$ -actions by diffeomorphisms of the manifold. In short, $\mathbb{R}$ -actions give rise to interesting $\mathbb{Z}$ -actions. In ergodic theory, a class of $\mathbb{Z}$ -actions are induced by $\mathbb{R}$ -actions using the method of *first returns*. This is studied via Kakutani equivalence and discussed in section “Kakutani Equivalence.”

Turning this around, $\mathbb{Z}$ -actions can induce $\mathbb{R}$ -actions using the method of *suspensions*. This technique allows lifting the complexity results from $\mathbb{Z}$ -actions to $\mathbb{R}$ -actions. The upshot is that $\mathbb{Z}$ and $\mathbb{R}$ -actions are closely related. For this reason Smale (1967) and others argue for studying $\mathbb{Z}$ -actions. Since the $\mathbb{Z}$ -actions are determined by their generator, this amounts to studying single transformations.

In summary, this entry will focus on $\mathbb{Z}$ -actions, which are determined by the generating transformation — in effect we are studying single transformations.

What structure and equivalence relations will be considered? While we will give examples of difficulties with structure theory in other contexts, the structure and classification theory of transformations described in this entry breaks very roughly into the *quantitative* theory and the *qualitative* theory. The ergodic theorem gives a framework for studying a function by repeated sampling along an orbit of a transformation T . As the number of samples grow, the averages converge to the average of the function over the whole space provided the given transformation is

ergodic. Hence, it can be viewed as the quantitative theory. The appropriate equivalence relation is conjugacy by measure-preserving transformations.

In (Smale 1967), Smale describes the study of transformations (in particular diffeomorphisms of manifolds) up to topological conjugacy as the *qualitative theory*. Symbolic shifts are also studied up to homeomorphisms (and up to their automorphism groups).

The understanding of the complexity of classifications in the quantitative theory is better developed than the understanding of the complexity of classifications in the qualitative theory. Ergodic theory classifications are discussed in section “Measure Isomorphism.” The nascent connections with the qualitative theory is discussed in section “Topological Conjugacy”, along with the many open problems.

To repeat, the two equivalence relations this survey will focus on are:

- Measure conjugacy
- Topological conjugacy

The first equivalence relation involves transformations that act on probability measure spaces X, Y , perhaps with other structural restrictions. (For example, volume preserving diffeomorphisms of a compact manifold.) Two such transformations S, T are *measure conjugate* (or measure isomorphic) if there is a measure isomorphism $\phi : X \rightarrow Y$ such that $S \circ \phi = \phi \circ T$ almost everywhere.

The second equivalence relation involves homeomorphisms acting on topological spaces X, Y , perhaps with other structural restrictions (S and T might be required to preserve smooth structures on X and Y .) A pair S, T are *topologically conjugate* (or topologically equivalent) if there is a homeomorphism $h : X \rightarrow Y$ such that $S \circ h = h \circ T$.

Choices of examples A theme of the survey is that many natural questions have intractable complexity, such as being infeasible using inherently countable information. For this to be most convincing, the classes should be taken to be as clear and concrete as possible in each context. In the context of the qualitative behavior of

diffeomorphisms, for example, this is the motivation for focusing on single transformations, taking the manifold to be as simple as possible (say the 2-torus) and the diffeomorphisms to be C^∞ . In many situations, the results are more general, but that is not the goal.

Homeomorphisms Often the class being given a structure theory consists of homeomorphisms with additional properties. In addition to smooth structure, one can consider minimal homeomorphisms, or topologically transitive homeomorphisms. An important example in section “Natural Classes That Are Not Borel” is the collection of topologically distal transformations.

We define the first two classes as they arise in many contexts. Let X be a topological space and $h : X \rightarrow X$ be a homeomorphism. Then h is *minimal* if every forward h -orbit of an $x \in X$ is dense. A weaker property is *topological transitivity*, which means that for some $x \in X$ the forward h -orbit of x is dense.

What a Classification isn't. To make sense of what a classification is, it might best be motivated by giving an example what a classification *isn't*. For concreteness, let us consider the space of measure-preserving transformations of the unit interval with the equivalence relation of measure isomorphism (conjugacy).

Because there are only continuum many measure-preserving transformations, there are at most continuum many equivalence classes $[T]$. By the Axiom of Choice, one can build a one-to-one function

$$F : \{[T] : T \in \text{MPT}\} \longrightarrow \mathbb{R}$$

where MPT is the collection of ergodic measure preserving transformations. By letting $C(T) = F([T])$, we get a map giving complete numerical invariants to the equivalence relation of measure isomorphism.

This is of course NOT what is meant by a classification, since simply evoking the Axiom of Choice gives no useful information whatsoever.

Requirements to be a structure theory or a classification A structure theory or a classification must be effective or computable in some sense.

There are at least three common versions of effectiveness:

- Computable *feasible* algorithms
- Computable using inherently finite techniques
- “Computable” using inherently countable techniques

We will say nothing about the first notion. Structure and Classification results that use inherently finite techniques are called *computable* or *recursive* classifications. This topic is well-covered in other literature (Braverman and Yampolsky 2009; Weihrauch 2000), and we are content with giving a couple of examples. Structure and Classification that uses inherently countable techniques are *Borel*. Techniques to show a classification is Borel, or not Borel, are the main topic of this entry.

We note that showing a class does *not* have a Borel structure or does *not* have a Borel classification is a stronger result than showing that there is no such computable theory. For this reason, this entry focusses on Borel classifications. There is much earlier work showing that various kinds of classifications are impossible using finitary techniques. These appear in places such as Mitin (1998); Ryzhikov (1985).

Benchmarking The complexity of the existing classifications can be measured by comparison with benchmark equivalence relations coming from outside dynamical systems. This benchmarking process can also give a rigorous method of saying that one classification is more complex than a second one. This benchmarking is called the theory of analytic equivalence relations and is discussed at length in section “Standard Mathematical Objects in Each Region.” A very brief introduction can also be found in Foreman (2018) (several of the diagrams in this entry were motivated by very similar diagrams in Foreman (2018)). There is an extensive literature about this topic, many of the results can be found in notes and books such as Gao (n.d.); Gao (2009); Hjorth (2000); Kechris (1995).

Examples of Structure and Classification Results in Dynamical Systems

To try to clarify the notions we begin by giving examples from dynamical systems (Much of this section is modeled on the introductory chapter of K. Peterson’s book (Peterson 1983).).

Hamiltonian Dynamics

These flows were developed by Hamilton to describe the evolution of physical systems (See Wikipedia [Hamiltonian system](#)). A Hamiltonian system is described by a twice differentiable (C^2) function $H(\mathbf{q}, \mathbf{p})$ from $\mathbb{R}^{6N}$ to $\mathbb{R}$, giving the energy of the system. The system is described by *Hamilton’s Equations*:

$$\begin{aligned}\frac{d\mathbf{p}}{dt} &= -\frac{\partial H}{\partial \mathbf{q}}, \\ \frac{d\mathbf{q}}{dt} &= +\frac{\partial H}{\partial \mathbf{p}}.\end{aligned}$$

The variables $\mathbf{p}, \mathbf{q} \in \mathbb{R}^{3N}$ are interpreted as the generalized position and momentum variables, and the solution $r(t)$ is viewed as the trajectory of a point in an initial position $r(0) \in \mathbb{R}^{6N}$. Putting this in the context of group actions, the solution is a one-parameter flow

$$T : \mathbb{R} \rightarrow \mathbb{R}^{6N}$$

What makes ergodic theory relevant to this example is *Liouville’s Theorem* that the Hamiltonian system $\{T_t\}_t$ preserves Lebesgue measure. For many fixed energies E the surface $H = E$ is a compact manifold and thus, on this manifold, the flow $\{T_t\}_t$ preserves an invariant probability measure (Khinchin 1949).

The Birkhoff ergodic theorem suggests the appropriate equivalence relation for studying systems with invariant measures:

Theorem 1. *Suppose that $(X, \mathcal{B}, \mu, T)$ is a measure-preserving system and $f \in L^1(\mu)$. Then the averages $(\frac{1}{N}) \sum_{n=0}^{N-1} f(T^n(x))$ converge almost everywhere to a T -invariant function $\bar{f}$ with*

$$\int \bar{f} d\mu = \int f d\mu.$$

In particular, if T is ergodic then the averages converge to the constant value $\int f d\mu$.

An informal interpretation of this theorem is that if one repeatedly samples an ergodic stationary system and then averages the results, the limit is the mean of the system. This suggests that the natural equivalence relation is *conjugacy by a measure-preserving transformation*, because this equivalence relation takes repeated sampling along trajectories to repeated sampling along trajectories. In different terminology, because *measure isomorphism* take averages to averages, it is the natural equivalence relation to study in this context.

Rotations of the Unit Circle

In this example, the class $\mathcal{C}$ is the collection of orientation preserving homeomorphisms of the unit circle. The structure is evident (as $\mathcal{C}$ itself forms the relevant Polish space X).

Let S^1 be the unit circle. It is convenient to view S^1 as the unit interval $[0, 1]$ with 0 and 1 identified. (An equivalent approach is to view S^1 as $[0, 2\pi]$ with the endpoints identified). Let $\pi : \mathbb{R} \rightarrow S^1$ be the map $x \mapsto [x]_1$ where $[x]_1$ is the positive fractional part of x .

If $f : S^1 \rightarrow S^1$ is an orientation preserving homeomorphism, then a *lift* of f is an increasing function $F : \mathbb{R} \rightarrow \mathbb{R}$ where $[F(x)] = h([x])$. Using the notion of lift we can define the rotation number of f (This discussion is largely based on (Barreira and Valls 2013), where the reader can see details).

Definition 2. Let

$$\rho(f) = \lim_{n \rightarrow \infty} \frac{F^n(x) - x}{n}. \quad (1)$$

Then $[\rho(f)]_1$ is called the rotation number of f

The following facts are necessary for this to make sense:

- $\rho(f)$ exists and is independent of x ,
- If ρ^* is defined by Eq. 1 using a different lift F^* of f , then for all x , $[\rho^*(f)]_1 = [\rho(f)]_1$.

Granting these facts, the rotation number is a well-defined member of $[0, 1)$ depending only on f . We will denote this number by $\rho(f)$ rather than $[\rho(f)]_1$.

The equivalence relation on homeomorphisms of the circle is conjugacy by a homeomorphism. So $f \sim g$ if there is a homeomorphism h such that

$$h \circ f = g \circ h.$$

If f is an arbitrary homeomorphism of the circle, then it is conjugate to a rotation preserving homeomorphism, so we focus on those.

One can check that:

Proposition 3. If f, g are orientation preserving homeomorphisms of S^1 that are conjugate by an orientation preserving homeomorphism, then $\rho(f) = \rho(g)$.

We will refer to statements like Proposition 3 as saying that $\rho(f)$ is a numerical invariant for the equivalence relation of orientation preserving topological conjugacy. Indeed having numerical invariants is one of the key notions of classifiability.

The classification question for ρ can then be phrased as:

If $\rho(f)$ a complete invariant?

Explicitly: can inequivalent f and g have $\rho(f) = \rho(g)$?

Clearly the rotation of the circle by $2\pi\theta$ radians, R_θ , has rotation number θ . So, the question of whether the rotation number is a complete invariant can be restated as asking whether $\rho(f) = \theta$ implies that f is conjugate to R_θ .

The following results show that $\rho(f)$ is a complete numerical invariant for functions with dense orbits. The first is due to Poincaré.

Theorem 4. (Poincaré) Suppose that f is a homeomorphism of S^1 with a dense orbit. Then f is topologically conjugate to R_θ for an irrational θ .

Denjoy showed that, if $f \in C^2$, then having an irrational rotation number implies having a dense orbit, however this is not true for general homeomorphisms (See (Wikipedia [Rotation number](#))).

Thus, homeomorphisms of the circle with a dense orbit is a story of a *successful* classification of the strongest kind: it is a classification by a numerical invariant. Moreover, Eq. 1 shows that with a sufficiently computable f , the numerical invariant can itself be computed and that the rotation number is a continuous function from X to $[0, 1)$. (See Katok and Hasselblatt 1995).

The next theorem gives information about the situation when f has a rational rotation number:

Theorem 5. *Let f be an orientation preserving homeomorphism from S^1 to S^1 . Then:*

1. *f has a rational rotation number if and only if f has a periodic point.*
2. *If $\rho(f) = p/q$ and $(p, q) = 1$ then f has periodic points and each periodic point has period q .*

For homeomorphisms with period points the issue relates to the classification of order preserving homeomorphisms of the unit interval, which is studied in Hjorth (2000).

After this entry was written Joint work of the author and Gorodetski shows that even for diffeomorphisms of S^1 , the equivalence relation of conjugacy-by-homeomorphisms turns out to be a maximal among equivalence classes reducible to S_∞ -actions and thus, quite complex. The complexity lies in diffeomorphisms with fixed points.

Symbolic Shifts

Let Σ be a finite or countable alphabet (Several authors refer to an alphabet as a “language,” a practice dating to Tarski). We let $\Sigma^{\mathbb{Z}}$ stand for the collection of bi-infinite countable sequences $\langle f(n) : n \in \mathbb{Z} \rangle$ written in the alphabet Σ .

The space $\Sigma^{\mathbb{Z}}$ carries the natural product topology. With respect to this topology, the shift map $sh : \Sigma^{\mathbb{Z}} \rightarrow \Sigma^{\mathbb{Z}}$ defined by:

$$sh(f)(n) = f(n + 1)$$

is a homeomorphism. A closed, shift-invariant $\mathbb{K} \subseteq \Sigma^{\mathbb{Z}}$ is called a *subshift*.

The structure and classification theories for symbolic shifts splits naturally into topological classifications and, when the symbolic shift has a specific measure associated with it, the classification up to measure isomorphism. The latter is discussed in the next section.

While many classes of symbolic shifts are studied, perhaps the most common are those shifts in a finite alphabet. Among the classes studied is the collection of *Shifts of Finite Type*. This is the class of subshifts that are determined by a fixed finite collection of words $w_1, w_2, \dots, w_k \in \Sigma^{<\mathbb{N}}$. Given this collection the subshift consists of all $f \in \Sigma^{\mathbb{Z}}$ such that for no interval is $f \upharpoonright [n, n + m)$ equal to some w_i .

It is straightforward to verify that the collection of subshifts of finite type is a countable subset of the compact subsets of $\Sigma^{\mathbb{Z}}$. Hence the structure aspect is easy to understand.

Each shift of finite type is determined by finite information (the forbidden words) and hence it makes sense to ask whether it is possible to determine whether two subshifts are topologically conjugate using inherently finite information. Various invariants exist which are derivable from *adjacency matrices* and two subshifts of finite type are conjugate by a homeomorphism if and only if their adjacency matrices are *strong shift equivalent*. This is an example of a *recursive reduction*. It reduces the equivalence relation of conjugacy by homeomorphism to the equivalence relation on the collection of adjacency matrices of being strong shift invariant (See section “What Is a Reduction?”).

However, the question of whether there is an algorithm to determine conjugacy by homeomorphisms for shifts of finite type remains an open problem (see Open Problem 2).

While other invariants exist for aperiodic subshifts such as *topological entropy*, these are not complete invariants for conjugacy by homeomorphism. In Camerlo and Gao (2001), it is shown that the classification of homeomorphisms of the Cantor set up to conjugacy has no Borel invariants – in fact it is maximal for S_∞ -actions (See

section “An Example of a Maximal S_∞ -Action from Dynamical Systems”).

In fact, Clemens (2009) was able to show that there are no Borel computable complete numerical invariants for conjugacy of subshifts. Clemens determined the exact Borel complexity of conjugacy of subshifts.

Measure Structure

There is an important and very well-developed study of transformations that preserve Borel measures that are not probability measures (Aaronson 1997), but to illustrate the connections with descriptive set theory, we consider probability measures.

We will discuss two types of questions about the structure of measure-preserving transformations: understanding a single transformation, and trying to understand the structure of factors and isomorphisms. In section “Measure Isomorphism”, we discuss the classification problem for measure-preserving systems and smooth measure-preserving systems. Kechris discusses more general results for countable amenable groups in (Kechris 2010).

A basic object in ergodic theory is $(X, \mathcal{B}, \mu, T)$ where $(X, \mathcal{B}, \mu)$ is a standard probability measure space and $T : X \rightarrow X$ is an invertible measure-preserving transformation. Following Furstenberg (1981), we will call these measure-preserving systems (and T a measure-preserving transformation) and reserve that term for probability measure-preserving systems. (We will only be considering invertible measure preserving systems so we drop the adjective *invertible*.)

While there are many models for the space of measure-preserving transformations, one is the collection of Lebesgue measure-preserving transformations of the unit interval equipped with Lebesgue measure. These are identified if they agree almost everywhere, and have the weak topology (the *Halmos topology*), which carries a complete separable metric (Halmos 1944, 1956). For a standard probability space, we will call the group of measure-preserving transformations $MPT(X)$. With the weak topology, it is a Polish group, with the operation of composition.

Two measure-preserving systems $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$ are (depending on the author) *isomorphic* or *conjugate* if there is an invertible measure-preserving transformation $\phi : X \rightarrow Y$ whose range has ν -measure one such that for μ -almost all $x \in X$

$$S \circ \phi(x) = \phi \circ T(x). \quad (2)$$

If ϕ is not invertible but still satisfies Eq. 2 almost-everywhere, then ϕ is a *factor map* and S is a factor of T .

If M is a compact metrizable space and $T : M \rightarrow M$ is a homeomorphism, then the Krylov-Bogoliubov Theorem shows that there is a T -invariant probability measure μ on the Borel subsets of M . The space P_T of T -invariant measures on M is convex and compact with the weak topology. Hence, the Krein-Milman theorem says that it is spanned by its extreme points (The interested reader can find explanatory discussions of these topics in Walters (1982) and Mirzakhani and Feng (2014)).

A measure-preserving transformation is *ergodic* if the only invariant measurable sets either have measure zero or measure one. The extreme points of the simplex of T -invariant measures on M are exactly the measure for which T is ergodic. (We will call these the *ergodic measures* when T is clear from context.) The Ergodic Decomposition Theorem says that if μ is a T -invariant measure then μ can be written as a direct integral of the ergodic measures. For the statement of the theorem, we let $\mathcal{M}(X)$ to be the space of invariant probability measures on X with the weak topology.

Theorem 6. *Let T be a measure-preserving transformation of a standard probability space $(X, \mathcal{B}, \mu)$. Then there is a measure space $(\Omega, \mathcal{C}, \nu)$ and a $\mathcal{C}$ -measurable map $\mu^* : \mathcal{C} \rightarrow \mathcal{M}(X)$ such that:*

- *For all $w \neq w'$, $\mu^*(w)$ and $\mu^*(w')$ are mutually singular and T -invariant.*
- *For all $Y \in \mathcal{B}$, $\mu(Y) = \int (\mu^*(w)(Y)) d\nu(w)$.*
- *For almost all $w \in \Omega$, $(X, \mathcal{B}, \mu^*(w), T)$ is ergodic.*

Moreover, the measure ν is unique up to sets of measure zero.

Thus, every invariant measure on M can be written as a direct integral of ergodic measures. We interpret this as saying that the ergodic measures are the building blocks of the invariant measures and focus the classification problem on the ergodic measures, or equivalently, on the ergodic measure-preserving transformations.

Notation We use the notation $EMPT(X)$ for the collection of ergodic measure-preserving transformations on $(X, \mathcal{B}, \mu)$.

The property of being ergodic is a conjugacy invariant so it is invariant under the conjugacy action of $MPT(X)$. Moreover, Halmos showed that the ergodic measure-preserving transformations are a dense $\mathcal{G}_\delta$ set in the weak topology on $MPT(X)$ (See, e.g., (Halmos 1956)).

Rolling Dice: Bernoulli Shifts

Two classes of measure-preserving systems provide very nice examples of both structure and a classification. One is the full shift on a finite alphabet with the Bernoulli measure. The other is Haar measure on a compact group. The latter is discussed in the next section.

Definition 7. Let Σ be a finite alphabet and ν be a measure on $(\Sigma, P(\Sigma))$. Let μ be the product measure $\nu^{\mathbb{Z}}$. Then μ is a shift-invariant measure on the full shift $\Sigma^{\mathbb{Z}}$. The system $(\Sigma^{\mathbb{Z}}, \mathcal{B}, \mu, sh)$ is called a Bernoulli Shift.

Informally: the coordinates of the Bernoulli shift are independent, identically distributed (iid). In particular the “time-zero” behavior is independent of the past.

Structure theory for Bernoulli shifts There is a successful structure theory for Bernoulli shifts. Ornstein showed (Ornstein and Weiss 1974) showed that being *finitely determined* is equivalent to being isomorphic to a Bernoulli shift and that being *very weak Bernoulli* implies being finitely determined. Ornstein and Weiss (1974) proved the converse: being finitely determined implies being very weak Bernoulli. Feldman (1974) used this characterization to show:

Theorem 8. (Feldman) The collection of measure-preserving transformations on the unit interval that are isomorphic to a Bernoulli shift on a finite alphabet is a Borel set.

Being isomorphic to a Bernoulli shift is an analytic condition, as we sketch in example 73. Feldman showed that the collection of very weak Bernoulli transformations is co-analytic. The theorem of Suslin (Corollary 72) says that a collection which is both analytic and co-analytic is Borel. Thus, the collection of Bernoulli shifts is Borel.

Classification theory for Bernoulli shifts Kolmogorov and Sinai, building on the work of Shannon defined a notion of *entropy* for measure-preserving transformations. Viewed as a function defined on the ergodic measure-preserving transformations, $EMPT$, *entropy* is a lower-semicontinuous. Kolmogorov showed that isomorphic Bernoulli shifts have the same entropy.

This sets the stage for the following result of Ornstein, which is exposited in (Ornstein 1970a, b, 1974):

Theorem 9. (Ornstein) Two Bernoulli shifts are isomorphic if and only if they have the same entropy.

Thus, the Bernoulli shifts are a paradigm of both a successful structure and a successful classification theory:

1. (Structure Theory) Bernoulli shifts can be identified in a Borel manner (they are the finitely determined transformations)
2. (Classification Theory) Entropy is a complete invariant for isomorphism.

Among the various differences between $\mathbb{Z}$ -actions and the actions of other groups is that the question of whether Bernoulli shifts based on F_2 -actions forms a Borel set among the F_2 -actions is an open problem. (See Open Problem 3)

A notion related to Bernoulli shifts is that of a Kolmogorov Automorphism, or for short, a $\mathcal{K}$ -automorphism. Informally: Bernoulli Shifts are characterized by the present being independent

from one moment earlier. $\mathcal{K}$ -automorphisms are characterized by the present being asymptotically independent of the past.

Definition 10. A $\mathcal{K}$ -automorphism is a measure-preserving transformation T of a standard probability space $(X, \mathcal{B}, \mu)$ such that there is a sub- σ -algebra $\mathcal{K} \subseteq \mathcal{B}$ with:

1. $\mathcal{K} \subseteq T\mathcal{K}$,
2. the union of the algebras $T^n\mathcal{K}$, for $n \geq 0$ generates $\mathcal{B}$ as a σ -algebra,
3. $\bigcap_{n=1}^{\infty} T^{-n}\mathcal{K} = \{\emptyset, X\}$.

There is a perfect set of non-isomorphic $\mathcal{K}$ -automorphisms, as was shown by Ornstein and Shields in (1973). In section “Reducing E_0 to Dynamical Systems”, it is shown that the very elegant classification theory for Bernoulli shifts does NOT extend to the $\mathcal{K}$ -automorphisms and little is known about classifying the $\mathcal{K}$ -automorphisms. For example, it is not known if the measure-isomorphism relation restricted to the $\mathcal{K}$ -automorphisms is Borel (See Open Problem 4.).

Symbolic Systems as Models for Measure-Preserving Transformations

Let $(X, \mathcal{B}, \mu)$ be a standard probability space, and $T : X \rightarrow X$ be an ergodic measure-preserving transformation. Then a set $\mathcal{A} \subseteq \mathcal{B}$ is a *generating set* if $\mathcal{B}$ is the smallest T -invariant σ -algebra containing $\mathcal{A}$ (where we identify sets if their symmetric difference is zero).

If $\mathcal{A}$ is a countable or finite generating set then we can make the elements of $\mathcal{A}$ disjoint and keep it a generating set. For this reason, the terminology *generating partitions* is frequently used. If $\mathcal{A} = \langle A_n : n \in \mathbb{N} \rangle$ is a partition of a set of measure one, then almost every x in X determines a $\mathbb{Z}$ -sequence of *names* $\langle A_{n_k} : k \in \mathbb{N} \rangle$ where $T^k(x) \in A_{n_k}$. The sequence is called the $\mathcal{A}$ -name of x . If $\mathcal{A}$ generates, then there is a set S of measure one such that for all $x \neq y \in S$ the $\mathcal{A}$ -name of x is different from the $\mathcal{A}$ -name of y .

Hence, on S there is a one-to-one map ϕ from X to $\mathcal{A}^{\mathbb{Z}}$ given by sending x to its $\mathcal{A}$ -name. Letting $\nu = \phi^*(\mu)$ (so copying the measure on X over to a measure on $\mathcal{A}^{\mathbb{Z}}$), we get an isomorphic copy of X as a shift space $(\mathcal{A}^{\mathbb{Z}}, \nu)$.

Rokhlin (1967) proved that there is always a countable generating set, and Krieger (1970, 1972a) proved that if the entropy of T is finite, then there is a finite generating set.

Krieger (1972b) showed the stronger result that if the original transformation T is ergodic and $K \subseteq \mathcal{A}^{\mathbb{Z}}$ is the support of ν , then (K, sh) is uniquely ergodic. We have outlined the proof of the following theorem

Theorem 11. Let $(X, \mathcal{B}, \mu)$ be an ergodic measure-preserving system. Then there is a countable alphabet $\mathcal{A}$ and a closed shift-invariant set $\mathbb{K} \subseteq \mathcal{A}^{\mathbb{Z}}$ carrying a unique shift-invariant measure ν such that $(\mathbb{K}, C, \nu, sh)$ is isomorphic to $(X, \mathcal{B}, \mu, T)$. If $(X, \mathcal{B}, \mu, T)$ has finite entropy then the alphabet $\mathcal{A}$ can be taken to be finite.

Koopman Operators

Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system. Define a map

$$U_T : L^2(X) \rightarrow L^2(X)$$

by setting $U_T([f]) = [f \circ T]$. Then U_T is a well-defined unitary operator on $L^2(X)$, and moreover, if S and T are isomorphic measure-preserving transformations, then U_S and U_T are unitarily equivalent.

Thus, it is possible to assign to each measure-preserving transformation a unitary operator in a way that sends conjugate transformations to unitarily equivalent operators. These operators are called the *Koopman Operators*. The equivalence classes of Koopman operators are thus a collection of invariants, but except for very special cases (such as in the next subsection) they are not complete invariants.

Translations on Compact Groups

We describe the *Halmos-von Neumann Theorem*. Let G be a compact group. Then G carries a Haar probability measure that is invariant under left multiplication. Suppose that $g \in G$ and Haar measure is ergodic for the map $T_g : G \rightarrow G$ given by $T_g(h) = gh$. Then G must be the closure of the powers of $g : G$ is a monothetic abelian group (The material presented in this example is explained in (Foreman 2000)).

An ergodic transformation $T: X \rightarrow X$ is said to have discrete spectrum if $L^2(U)$ is spanned by the eigenfunctions of U_T . The eigenfunctions can be taken to have constant modulus one and are closed under multiplication. Since T is measure-preserving the eigenvalues have modulus one and form a subgroup of the unit circle. Hence, by Pontryagin duality their dual is a compact monothetic abelian group G . The identity map from the set of eigenvalues into the unit circle is a generator of this group. If g is the identity character and T_g is translation by g , then there is a unitary operator conjugating U_{T_g} with U_T that preserves multiplication of bounded functions in $L^2(G)$. From this one can compute a measure isomorphism between (G, T_g) and (X, T) .

Again by Pontryagin duality, if G is a compact group, then $L^2(G)$ is spanned by the characters of G . An immediate calculation shows that if G is monothetic, with generator g and T is translation by g , then the characters of G are eigenvalues – hence T has discrete spectrum.

The Halmos-von Neumann theorem (Halmos and von Neumann 1942) gives both a structure theorem and a classification theorem.

Theorem 12. (Halmos-von Neumann) *Let T be an ergodic transformation. Then:*

1. (Structure) *T has discrete spectrum if and only if it is isomorphic to an ergodic translation on a compact group with Haar measure.*
2. (Classification) *If S, T have discrete spectrum, then S and T are conjugate by a measure-preserving transformation if and only if the Koopman operators associated to S and T have the same eigenvalues.*

For a proof of this theorem and more on this topic see Walters (1982).

Presentations

Often transformations with various relevant properties one wants to understand can be presented in different contexts. The context in which they are presented can be important. For example, two isomorphic measure-preserving transformations,

presented with measures on different spaces can have different simplexes of invariant measures. Another example is Theorem 11 that shows that every measure-preserving transformation can be presented as a symbolic shift.

Striking examples include Bernoulli Shifts: there are Bernoulli shifts that are presented as real-analytic measure-preserving transformations of the torus (with the Poulsen simplex of invariant measures). Bernoulli shifts can also be presented as uniquely ergodic symbolic systems. The issue of presentation is quite different from the issue of isomorphism, although topological conjugacy is quite sensitive to presentation.

Presentations *can* affect intrinsic measure theoretic properties. For example, Kushnirenko showed that a diffeomorphism of a compact manifold preserving a smooth volume element has finite entropy. One of the most prominent open problems in the area is the converse: can every finite entropy measure-preserving transformation be presented as a C^∞ transformation on a compact manifold that preserves a volume form? (See Open Problem 8.)

Smooth Dynamics

In the context of this entry, *smooth dynamics* will mean studying diffeomorphisms of compact manifolds. The equivalence relation is that of conjugacy by homeomorphisms. We repeat the definitions from the introduction.

Explicitly, we let M be a smooth manifold, usually taken to be compact. We will be given a class C of diffeomorphisms of M described with a structure theory. Two elements $f, g \in C$ will be taken to be *equivalent* if there is a homeomorphism $h: M \rightarrow M$ such that

$$f \circ h = h \circ g. \quad (3)$$

Smale (1967) calls this equivalence relation the *qualitative theory* of diffeomorphisms. He says that “the same phenomena and problems of the qualitative theory of ordinary differential equations [are] present in the simplest form in the diffeomorphism problem.” He argues that we should study the equivalence relation given when the conjugacy map h is a homeomorphism,

rather than being a diffeomorphism, because “differential conjugacy is too fine” to reveal the relevant properties such as structural stability.

Some examples of classes of diffeomorphisms There are many well-studied classes of diffeomorphisms. For the purposes of this entry, we will assume that they are C^∞ and usually assume that the manifolds M are compact. We start with two of the well-known classes of diffeomorphisms.

Suppose $f: M \rightarrow M$ is a diffeomorphism and that $\Lambda \subseteq M$ is invariant. Then Λ is *hyperbolic* for f if there are constants $C > 0$ and $0 < \lambda < 1$ and a decomposition of the tangent bundle restricted to Λ as:

$$\begin{aligned} TM_\Lambda &= E^s \oplus E^u \\ Tf(E^u) &= E^u \\ Tf(E^s) &= E^s \end{aligned}$$

and for $v \in E^s$, $n > 0$

$$\|Tf^n(x)v\| \leq C\lambda^n \|v\|$$

and for $v \in E^u$, $n \geq 0$

$$\|Tf^{-n}(x)v\| \leq C\lambda^n \|v\|$$

Definition 13. A diffeomorphism $f: M \rightarrow M$ is Anosov if the whole manifold M is hyperbolic for f .

Suppose that x is a periodic point of order m and $x, f(x), \dots, f^{m-1}(x)$ are hyperbolic points. Then the stable and unstable manifolds of x are:

$$\begin{aligned} W^s &= \{y \in M \mid f^{mk}(y) \rightarrow x \text{ as } k \rightarrow \infty\} \\ W^u &= \{y \in M \mid f^{-mk}(y) \rightarrow x \text{ as } k \rightarrow \infty\} \end{aligned}$$

These are immersed Euclidean spaces of dimension corresponding to the number of eigenvalues of the differential $Df^m(x)$ that are greater or less than one in absolute value.

Definition 14. A diffeomorphism $f: M \rightarrow M$ is Morse-Smale if and only if there are a finite collection of periodic orbits $P_1, \dots, P_l$ such that:

1. P_i is a hyperbolic periodic point for $i = 1, \dots, l$,
2. $\bigcup_{i=1}^l W^s(P_i) = M$,
3. $\bigcup_{i=1}^l W^u(P_i) = M$,
4. $W^u(P_i)$ and $W^s(P_j)$ are transversal for all i, j .

1. (Structure) Both the Anosov and the Morse-Smale diffeomorphisms are examples of *structurally stable* diffeomorphisms. The structurally stable diffeomorphisms are those f for which there is a C^1 -neighborhood U such that every $g \in U$ is topologically conjugate to f . Thus

- (a) The Anosov diffeomorphisms form an open set
- (b) The Morse-Smale diffeomorphisms form an open set
- (c) The structurally stable diffeomorphisms form an open set

It follows that there are only countably many classes of Anosov and Morse-Smale diffeomorphisms.

It would be a nice picture if the collection of structurally stable diffeomorphisms formed an open and *dense* subset of the diffeomorphisms, but this was refuted by Newhouse (Newhouse 1970, 1974).

2. (Classification) Each of the three classes form an open subset of the space of diffeomorphisms. Each of the open sets decompose into a union of open subsets corresponding to each of the equivalence classes. Thus for each class, there is a single countable list of representatives and radii that capture all of the members of the class. The representatives can be taken to be rational (in the appropriate sense) and the radii can be taken to be of the form $1/n$. Hence the countable lists can be viewed as members of a Polish space.

The upshot is that for each of the three classes, one can assign complete numerical invariants in a Borel (even continuous) way. The equivalence relation is Borel reducible to the equality relation on a Polish space.

In dimension 2, more intuitive invariants can be assigned in a computable way. For Anosov

diffeomorphisms on the 2-torus these are hyperbolic elements of $SL_2(\mathbb{Z})$ and for Morse-Smale diffeomorphisms, they are graphs with colored edges (Oshemkov and Sharko 1998; Peixoto 1973). Similar results hold for Morse-Smale diffeomorphisms in dimension 3 (Bonatti et al. 2019).

An alert reader will notice a problem with both the structure and classification results just stated. The program is to classify equivalence classes of diffeomorphisms, and the definition of Anosov, Morse-Smale, and structurally stable diffeomorphisms are not invariant under topological conjugacy. Ideally, given a diffeomorphism f that is topologically conjugate to an element of one of these classes, there would be a way of constructing a g equivalent to f that actually belongs to the class. This seems to be an open problem (See Open Problem 9).

More Detailed Structure Theory

So far the term “structure theory” has been used solely for the purpose of determining membership in a class. However, in several cases a structure theory gives more information – for example, about the factor structure of a member of a class. The Furstenberg structure theorem for topologically distal transformations and the Furstenberg-Zimmer theory for measure-preserving transformations are examples of this. These are discussed in section “Natural Classes That Are Not Borel.”

Summary

The previous sections are intended to give a context for *structure theory* and *classification theory*. These are somewhat vague terms, but the intention is the following:

- Structure theory takes place in the context of an ambient Polish space and gives a method for determining membership in a class C . It can also give information about the factors of a transformation or explicit information about the way the transformation moves elements in the space it acts on.
- Classification theory considers a class C and an equivalence relation E on that class. It is concerned with determining when two elements of C are E -equivalent.

A structure theorem is concerned with a single transformation, and classification theory is concerned with pairs of transformations.

Descriptive Complexity

For a structure theory or classification theory to be useful, it is important that it be computable in some sense. As the example given in section “Introduction: What Is a Dynamical System? What Is Structure? What Is a Classification?” that used the Axiom of Choice illustrated, classifications or structure theorems that are not in some sense computable are likely to be meaningless.

The structure theoretical levels of complexity are easier to describe than the *complexity benchmarks* for classification, as it is a series of yes-no questions. Recall from the discussion in the introduction the three basic questions about how effective a structure theory can be are:

1. Is the structure theory computable with realistic resources assumptions?
2. Can it be carried out with inherently finite information?
3. Can it be carried out with inherently countable information?

We will ignore the first question and say little about the second. The main point of the entry is that there are natural examples of dynamical systems fail even the third question. The third question can be rephrased as asking whether the structure theory shows that the class C is Borel (Of course, once a structure theory has been shown to be Borel, the question arise to determine what level in the Borel hierarchy it lies in. There is such a literature, but it is not addressed in this survey.).

Remark. *In this survey, we only consider equivalence relations that are analytic. These include all orbit equivalence relations of Polish groups. (See example 73.) Some natural equivalence relations are not analytic (such the equivalence relation on distal flows of having the same distal height). The theory is not nearly as well developed for these so they are not covered in this survey.*

What Is a Reduction?

There are many jokes about mathematicians where the punchline is that the way they solved problem A was to “reduce it to a problem B they already knew how to solve.” (A google search in 2018 gave over 115 million hits to the search “reduce it to a problem you can solve joke.”)

The idea behind the joke is that to solve a problem A you (somewhat constructively) reduce it to a problem B you already know how to solve. One can state this mathematically in the following way.

Reductions of sets

Let X and Y be Polish spaces and $A \subseteq X$, $B \subseteq Y$. Then A is

_____ - reducible

to B provided that there is a function $f: X \rightarrow Y$ such that for all $x \in X$

$$x \in A \text{ if and only if } f(x) \in B.$$

Fill in the blank As the example given earlier showed, without some restrictions on the function f this notion is not meaningful. If B and $Y \setminus B$ are non-empty then every set $A \subseteq X$ is reducible to B : fix $b_0 \in B$ and $b_1 \in Y \setminus B$ and define $f(x) = b_0$ if $x \in A$ and $f(x) = b_1$ if $x \in X \setminus A$. Then f reduces A to B . Thus, one must fill in the blank with some condition on the concreteness of f .

In the complexity theory world of computer science, the blank is often filled in by some statement “linearly,” “polynomially,” “exponentially,” or “recursively.” In the context of this entry, we focus on Borel functions f .

Starting in the 1950s, in the subject then known as Recursion Theory (Soare 1987) and now called Computability Theory, the joke was reversed:

To show a problem B is not solvable, you find an unsolvable problem A and reduce it to B by a _____ function.

If B were solvable, and A is reduced to B , A would also be solvable, yielding a contradiction. To make the definition complete, one has to fill in the blank, usually from one of the three types of functions: feasibly computable, recursively computable or Borel.

Indeed this has a long history: the unsolvability of the word problem for finitely presented groups was shown by reducing an incomputable set to the word problem using a computable function.

The focus of this survey is the notion of *Borel Reducibility* (This idea originated in the late 1980s in the work of Friedman and Stanley (1989) related to model theory and independently Harrington, Kechris, and Louveau who used it in the context of analysis.), with some very short excursions into continuous reducibility. To repeat the official definition, we let X and Y be Polish spaces. A set

$$A \subseteq X \text{ is Borel reducible to } B \subseteq Y$$

if and only if there is a Borel measurable function $f: X \rightarrow Y$ such that for all $x \in X$

$$x \in A \text{ if and only if } f(x) \in B.$$

Thus the inverse image of B under f is A . Because the definition of *Borel measurable* implies that the inverse image of a Borel set is Borel, it follows that

if A is not Borel, then B is not Borel.

Notation: We write $A \leqslant_B^1 B$ if A is Borel reducible to B . In this notation, the previous remarks can be written for the record as:

Remark 15. Let X, Y be Polish spaces. Let $A \subseteq X$ and $B \subseteq Y$, with B Borel. If $A \leqslant_B^1 B$ then A is Borel.

Reductions of Equivalence Relations

For the classification problem a two dimensional version is more useful. Let X and Y be Polish spaces and $E \subseteq X \times X$, $F \subseteq Y \times Y$ be equivalence relations. Then

$$E \text{ is Borel reducible to } F$$

if and only if there is a (unary) Borel function $f: X \rightarrow Y$ such that for all $x_1, x_2 \in X$

$x_1 E x_2$ if and only if $f(x_1) F f(x_2)$.

We use the notation $E \leqslant_{\mathcal{B}}^2 F$.

It is important to recognize how the two definitions differ. In the second, two-dimensional definition, f has domain X , not $X \times X$. So, in essence:

f assigns an F -equivalence class to each x in X in a manner that two elements of X get assigned the same F -equivalence class if and only if they are in the same E -equivalence class.

Viewing the classes of F as invariants, this is paraphrasing of the statement that f computes the invariants assigned to members of X .

Thus for $\leqslant_{\mathcal{B}}^2$, f is a function that assigns values to pairs (x_1, x_2) by considering each x_i separately. However viewed as a function from the Polish space $W = (X \times X)$ to the Polish space $Z = (Y \times Y)$ it is also a one-dimensional reduction.

Remark 16. Suppose that $E \subseteq X \times X$ and $F \subseteq Y \times Y$, then $E \leqslant_{\mathcal{B}}^2 F$ implies $E \leqslant_{\mathcal{B}}^1 F$.

Continuous Reductions

An important special case of Borel reducibility is when the function f is continuous. This is stronger than being Borel reducible. Essentially, every general statement below applies to continuous reductions as well as Borel reductions. If A is continuously reducible to B we write $A \leqslant_C^1 B$ and if E is continuously reducible to F we write $E \leqslant_C^2 F$. If continuity is clear from context we omit the subscript C and if the number of variables is clear from context we will omit the superscript in both cases.

The Pre-ordering

For both Borel and continuous reductions and both the one- and two-dimensional cases, the relation of $\leqslant_{\mathcal{B}}$ and $\leqslant_C$ is transitive. To see this we note that if $f: X \rightarrow Y$ is a reduction (in the appropriate sense) and $g: Y \rightarrow Z$ is a reduction (in the same sense), then $g \circ f$ is a reduction in the same sense.

As with all pre-orderings, we can have *bi-reducible sets* and bi-reducible equivalence relations. There are examples where $A \leqslant_{\mathcal{B}} B$ and $B \leqslant_{\mathcal{B}} A$ but there is no bijection $f: X \rightarrow Y$ such that f reduces A to B , and f^{-1} reduces B to A . (The

analogue to the Cantor-Schroeder-Bernstein theorem fails.) We use the notation $A \sim_{\mathcal{B}} B$ (or $\sim_C$ or...) to mean the equivalence relation of bi-reducibility: $A \leqslant_{\mathcal{B}} B$ and $B \leqslant_{\mathcal{B}} A$.

Let $\sim$ be the relation $A \leqslant_{\mathcal{B}} B$ and $B \leqslant_{\mathcal{B}} A$. Then $\leqslant_{\mathcal{B}}$ gives a partial ordering of $\sim$ – classes of sets or equivalence relations depending on the dimension one is working in.

Caveat! There are several diagrams in this entry that give the relationship of reducibility between various classes of objects. The regions signify structures that are reducible to the type of object used to label the region. So, for example, an object is in the region Polish group actions if it can be reduced to a *Polish group action*. In particular, this applies to Figs. 2, 4, 5, and 10. In the diagrams, a line segment going upwards means *Borel Reducible* and a pointed arrow means that the downwards reduction does not hold. Not having an arrow means that bi-reducibility is open. An arrow at both ends indicates that bi-reducibility does hold.

Universality Given a class C of either sets (in the one-dimensional case) or equivalence relations (in the two dimensional case) an object $B \in C$ is *universal* for C if every $A \in C$ is reducible to it. The interpretation of this is that B is the most complex element of C . This is used for all of the notions: Borel reductions, continuous reductions, and computable reductions.

Terminology warning: *Maximal* is often used as a synonym of *universal*. Less frequently *complete* is also used. We will use *maximal* when talking about the $\leqslant_{\mathcal{B}}^2$ relation on equivalence relations, the two dimensional case. We will use *complete* for $\leqslant_{\mathcal{B}}^1$ restricted to sets, the one-dimensional case.

Because of the transitivity of the notions of reducibility, if $B, C \in C$ and B is universal for C and $B \leqslant_{\mathcal{B}} C$, then C is universal for C . It follows that all universal sets are bi-reducible to each other.

The upshot We can view the notions $\leqslant_{\mathcal{B}}$ and $\leqslant_C$ as measures of complexity. If $A \leqslant_{\mathcal{B}} B$, then B is *at least as complex* as A . In a given class C there may be $\leqslant$ -maximal elements. They represent the equivalence class of the most complex elements of C .

Reducing Ill-Founded Trees

There is a basic method of establishing a set is *not* Borel. Because it arises in the proofs of several theorems, we include it here rather than in the appendix A. It gives a canonical example of a complete analytic set that is combinatorially convenient to work with.

The space of Trees For applications, a common complete analytic (and so non-Borel), set is the collection of ill-founded trees. Let $\mathbb{N}^{<\mathbb{N}}$ be the collection of finite sequences of natural numbers, viewed as being of the form $\sigma = \langle n_0, n_1, \dots, n_{l-1} \rangle$. The number l is the length of σ . The Polish space of trees is defined as follows:

1. A *tree* is a set $\mathcal{T} \subseteq \mathbb{N}^{<\mathbb{N}}$ that is closed under initial segments. So if $\sigma = \langle n_0, n_1, \dots, n_{l-1} \rangle \in \mathcal{T}$ and $k < l$ then $\tau = \langle n_0, n_1, \dots, n_{k-1} \rangle \in \mathcal{T}$. (It is important to note that the trees are not required to be finitely branching. Every tree has the empty sequence in it.)
2. The space *Trees* is the collection of trees $\mathcal{T} \subseteq \mathbb{N}^{<\mathbb{N}}$.
3. A basic open set in the topology on the space of trees is given by the finite sequence σ and $\langle \sigma \rangle = \{ \mathcal{T} : \sigma \in \mathcal{T} \}$.

Recasting the definition, putting the discrete topology on $\{0, 1\}$, and the product topology on $\{0, 1\}^{\mathbb{N}^{<\mathbb{N}}}$ gives a Polish space X homeomorphic to the Cantor set. Each tree $\mathcal{T}$ can be identified with its characteristic function $X_{\mathcal{T}} : \mathbb{N}^{<\mathbb{N}} \rightarrow \{0, 1\}$, and hence, an element of X . It is not difficult to verify that the induced topology on $\{\chi_{\mathcal{T}} : \mathcal{T} \in \text{Trees}\} \subseteq X$ gives a space homeomorphic to *Trees*.

The first condition in the definition of a tree gives a countable sequence of requirements corresponding to open sets, showing that $\{\chi_{\mathcal{T}} : \mathcal{T} \in \text{Trees}\}$ is a closed subset of X and hence a Polish space.

Definition 17. Let $\mathcal{T} \in \text{Trees}$ and $f : \mathbb{N} \rightarrow \mathbb{N}$. Then f is an *infinite branch through* $\mathcal{T}$ if for all $l \in \mathbb{N}$ the sequence $\sigma = \langle f(0), f(1), \dots, f(l-1) \rangle \in \mathcal{T}$.

The König infinity lemma implies that every infinite, finitely branching tree has an infinite

branch. However, in this context the trees in *Trees* are allowed to be infinitely branching.

Remark. In alternate language, not used in descriptive set theory, the space of trees is the space of rooted, labeled acyclic graphs (allowing countably infinite valence), and the collection of ill-founded trees are those graphs with non-trivial ends.

Put the discrete topology on $\mathbb{N}$ and consider $\mathbb{N}^{\mathbb{N}}$ with the product topology. This yields a Polish space called the *Baire space*. Then, $B = \{(\mathcal{T}, f) : f \in \mathbb{N}^{\mathbb{N}}, \mathcal{T} \in \text{Trees} \text{ and } f \text{ is a branch through } \mathcal{T}\}$ is a closed subset of $\text{Trees} \times \mathbb{N}^{\mathbb{N}}$. Because

$$\begin{aligned} \{\text{ill-founded trees}\} \\ = \{\mathcal{T} : \text{for some } f \in \mathbb{N}^{\mathbb{N}} (\mathcal{T}, f) \in B\}, \end{aligned}$$

it follows that the collection of ill-founded trees is an analytic subset of *Trees*.

A tree is *well-founded* if it is not ill-founded. Hence, the well-founded trees are the complement of the ill-founded trees and thus *co-analytic*.

Fundamental results about the collection of ill-founded trees include:

Theorem 18. Let *Trees* be the space of trees and $IFT \subseteq \text{Trees}$ be the collection of ill-founded trees. Then:

1. If X is an arbitrary Polish space and $A \subseteq X$ is analytic then A is Borel reducible to IFT .
2. IFT is an analytic set that is not Borel.

Corollary 19. Let $WF \subseteq \text{Trees}$ be the collection of well-founded trees. Then

1. If X is an arbitrary Polish space and $C \subseteq X$ is co-analytic, then C is Borel reducible to WF .
2. WF is a co-analytic set that is not Borel.

In the language of section “What Is a Reduction?”, the collection of ill-founded trees, as a subset of *Trees*, is a complete analytic set. It also follows from Theorem 18 that there are non-Borel analytic sets.

Basic Technique for showing a set is analytic or co-analytic but not Borel. The basic technique for showing that a set C is not Borel is reducing a known nonBorel set to C . To use this effectively, one needs a starting point. The sets *IFT* or *WF* play this role for analytic and co-analytic sets.

Let X be a Polish space and $C \subseteq X$ be analytic. Then to show that C is complete analytic, and so not Borel, it suffices to build a reduction

$$f : \text{Trees} \rightarrow X$$

such that

$$\mathcal{T} \text{ is ill-founded if and only if } f(\mathcal{T}) \in C.$$

By Theorem 18, it follows that C is not Borel.

Similarly the basic technique for showing that a co-analytic set C is complete co-analytic is to build a reduction

$$f : \text{Trees} \rightarrow X$$

such that

$$\mathcal{T} \text{ is well-founded if and only if } f(\mathcal{T}) \in C.$$

The most basic Benchmarks The first collection of benchmarks is determining whether a classification is computable, Borel or analytic-and-not-Borel. These are the basic levels of descriptive complexity. If an example is Borel, one can also ask what level it lives in the Borel hierarchy.

Paraphrasing this informally, the question is whether every classification can be done using inherently finite techniques, inherently countable techniques, or whether it is impossible to do the classification using inherently countable techniques. If a classification is complete analytic or co-analytic, then there is no possible classification that is feasible with countable resources.

Complexity in Structure Theory

For our purposes, structure theories for classes $C \subset X$ are concerned with two questions:

- Can you determine whether an arbitrary $x \in X$ belongs to C ?
- How complicated is a given method for building elements of C ?

The complexity of the class The basic question is what level of descriptive complexity the class lies in relative to the basic benchmarks, given above.

An early example: weak mixing vs strong mixing The first example of the use of descriptive set theory to solve a natural structure problem in dynamical systems is due to Halmos and Rokhlin. A prominent open problem in the 1940s was whether the property of a transformation being *weakly mixing* implied the property of being *strongly mixing*.

In 1944, in a paper called *In general a measure-preserving transformation is mixing* (Halmos 1944), Halmos proved that the collection of weakly mixing transformations is a dense G_δ set in the weak topology. In a paper published in 1948 with a title translated as *A general measure-preserving transformation is not mixing* (Rohlin 1948), Rokhlin showed that the strong mixing transformations is meager (first category). The combination of the two papers show that there are weakly mixing transformations that are not strongly mixing.

A proof that explicitly uses complexity shows that the collection of strongly mixing transformations is not Π_2^0 , and hence cannot be the same class as the weakly mixing transformations. (In fact the class is Π_3^0 .) In particular, there is a weakly mixing transformation that is not strongly mixing, as it is complete Σ_2^0 .

An explicit example of a weakly mixing transformation that is not strongly mixing was given using Gaussian systems was provided by Maruyama in 1949 (Theorem 11 of (Maruyama 1949)). A very natural example of such a transformation is due to Chacon (1969).

Examples at Three Levels of Complexity

We now give examples of each level beyond recursively computable.

Non-recursive Classes

Example. (Braverman, Yampolsky) (Braverman and Yampolsky 2009) *There is a computable complex number s such that membership in the Julia set associated with*

$$f(z) = z^2 + s$$

is not computable.

Borel Classes

Borel Classes of measure-preserving transformations As mentioned above, the weakly mixing and strongly mixing classes of measure-preserving transformations of the unit interval are Borel. The multiple mixing counterparts are as well.

In the topological category, the collections of uniformly continuous maps from the open unit interval to the open unit interval are easily seen to be Π_2^0 in the space of continuous transformations.

Indeed most standard classes of transformations are Borel. As explained in the discussion above the collection of Bernoulli shifts is Borel. However, proving this uses Suslin's Theorem, which involves analytic sets.

Other well-known classes of measure-preserving transformations are Borel and the proof uses a significant amount of work. For example, the class of rank-one-rigid transformations (King 1986).

Borel Classes of homeomorphisms of compact spaces A potential problem with this collection is that there are a variety of natural spaces in which to set the various questions. For example, every compact $\mathbb{Z}$ -flow can be viewed as a compact subset of the $\mathbb{H}^{\mathbb{Z}}$ or as a compact subset of $\mathbb{H}^2$ (where $\mathbb{H}$ is the Hilbert cube) or as an appropriate subset of $\mathbb{N}^{\mathbb{N}}$. All of the known natural codings of the collection of homeomorphisms of compact metric spaces agree on what collections of homeomorphisms are Borel. (This is discussed in (Beleznay and Foreman 1995).)

Examples include:

- $\{(X, T): T \text{ is topologically transitive and } X \text{ is Polish}\}$

- $\{(X, T): T \text{ is minimal and } X \text{ is Polish}\}$

Proofs of these results (and later calculations of an upper bound on the complexity of topologically distal transformations) were novel in the early 1990s and appear in Beleznay and Foreman (1995) (Lemma 3.3). The second item is due to Kechris.

Borel Classes of diffeomorphisms We have already seen that the collection of structurally stable diffeomorphisms is Borel – in fact it is open as are the collections of Anosov and Morse-Smale diffeomorphisms.

To give a flavor of tools used for more complicated classes, we use the following results in Foreman and Gorodetski (2022):

Denote by $\text{Fix}(f)$, the set of fixed points of the map $f: M \rightarrow M$, i.e.,

$$\text{Fix}(f) = \{x \in M \mid f(x) = x\},$$

and by $\text{Per}_n(f)$ the set of periodic points of period n , i.e.,

$$\text{Per}_n(f) = \left\{x \in M \mid f^n(x) = x \text{ and for all } m < n, f^m(x) \neq x\right\}.$$

Proposition. *Let M be a compact manifold. For any $r = 0, 1, 2, \dots, \infty$ the following statements hold:*

1. *The set $\mathcal{U}_0 = \{f \in \text{Diff}^r(M) : \text{Fix}(f) = \emptyset\}$ is open.*
2. *For each $m \in \mathbb{N}$ the set $\mathcal{U}_m = \{f \in \text{Diff}^r(M) : |\text{Fix}(f)| = m\}$ is Borel.*
3. *For each $m, n \in \mathbb{N}$ the set $\mathcal{U}_m^n = \{f \in \text{Diff}^r(M) : |\text{Per}_n(f)| = m\}$ is Borel.*
4. *For any $A: \mathbb{N} \rightarrow \mathbb{N}$ the set*

$$\mathcal{U}_A = \left\{f \in \text{Diff}^r(M) : |\text{Per}_n(f)| = A(n), n = 0, 1, 2, \dots\right\}$$

is Borel.

5. The set

$$\left\{ (A, f) : |Per_n(f)| = A(n), n = 0, 1, 2, \dots, \right. \\ \left. \text{where } A : \mathbb{N} \rightarrow \mathbb{N} \right\}$$

in the space $\mathbb{N}^{\mathbb{N}} \times \text{Diff}^r(M)$ is Borel.

We give the proof that appeared in Foreman and Gorodetski (2022) for items 1–3. The first item is immediate: since M is compact there is a minimum distance between x and $f(x)$. The set of diffeomorphisms whose minimum distance is bigger than $1/n$ is open, hence the set of diffeomorphisms with no fixed points is open.

To see the second item: To see that $\mathcal{U}_1$ is Borel, let us choose a countable base for the topology on M , $\{V_k\}_{k \in \mathbb{N}}$, such that

$$\text{diam } V_1 \geq \text{diam } V_2 \geq \text{diam } V_3 \geq \dots$$

and $\text{diam } V_k \rightarrow 0$ as $k \rightarrow \infty$. The set

$$\mathcal{V}_k = \{f \in \text{Diff}^r(M) \mid f \text{ has no fixed points outside of } V_k\}$$

is open, and hence the set $\mathcal{U}_1 = (\bigcap_{K \geq 1} \bigcup_{k=K}^{\infty} \mathcal{V}_k) \setminus \mathcal{U}_0$ is Borel.

To see that $\mathcal{U}_2$ is Borel, notice that each of the sets

$$\mathcal{V}_{k_1, k_2} = \{f \in \text{Diff}^r(M) \mid f \text{ has no fixed points outside of } V_{k_1} \cup V_{k_2}\}.$$

is open, and hence

$$\mathcal{U}_2 = \left(\bigcap_{K \geq 1} \bigcup_{k_1, k_2 \geq K} \mathcal{V}_{k_1, k_2} \right) \setminus (\mathcal{U}_0 \cup \mathcal{U}_1)$$

is Borel. It is clear how to continue for n fixed points by induction.

Example. Artin-Mazur Diffeomorphisms It follows easily from the proposition that the function giving the exponential rate of growth of the number of periodic points $G : \text{Diff}^r(M) \rightarrow \mathbb{R} \cup \{\pm\infty\}$ defined by

$$G(f) = \limsup_{n \rightarrow \infty} \frac{1}{n} \log |Per_n(f)|$$

is Borel. In particular, the set of Artin-Mazur diffeomorphisms (defined as $\{f : G(f) < \infty\}$) is Borel.

Natural Classes that Are Not Borel

In the topological setting A classical collection, due to Hilbert (see Zippin 2013) is defined as follows.

Definition 20. Let (X, d) be a metric space and $T : X \rightarrow X$ be a homeomorphism. Then (X, T) is (topologically) distal if for all distinct $x, y \in X$, $\inf_{n \in \mathbb{Z}} \{d(T^n x, T^n y)\} > 0$.

Clearly, isometries are examples of distal transformations. Moreover, the distal transformations are closed under taking skew products with compact metric groups. Using these two facts one can generate a wide class of distal transformations.

Fact. The collection of distal transformations is a co-analytic set. This is exposted in example 74.

Definition 21. Let (X, d, T) and (Y, d', S) be a compact separable metric spaces and $\pi : X \rightarrow Y$ be a factor map. Then X is an isometric extension of Y if there is a real valued function $\rho(x_1, x_2)$ defined for all pairs (x_1, x_2) in the same fibre of X (i.e., whenever $\pi(x_1) = \pi(x_2)$) and such that

1. $\rho(x_1, x_2)$ is continuous as a function on the subset of $X \times X$ defined by the condition $\pi(x_1) = \pi(x_2)$.
2. For each $y \in Y$, $\rho(x_1, x_2)$ defines a metric on the fiber $\pi^{-1}(y)$ under which this fiber is isometric to a fixed homogeneous metric space.
3. $\rho(Tx_1, Tx_2) = \rho(x_1, x_2)$ for all x_1, x_2 in the same fibre of X over Y .

The next definition captures the process of iterating the isometric extensions transfinitely:

Definition 22. (X, T) is a quasi-isometric extension of (Y, T) if (Y, T) is a factor of (X, T) , and there is an ordinal η and a factor (X_ξ, T) of (X, T) for each $\xi \leq \eta$ such that:

1. $(X_0, T) = (Y, T); (X_\eta, T) = (X, T)$.
2. if $\xi < \xi'$ then (X_ξ, T) is a factor of $(X_{\xi'}, T)$.
3. $(X_{\xi+1}, T)$ is an isometric extension of (X_ξ, T) .
4. if ξ is a limit ordinal then (X_ξ, T) is the inverse limit of $\langle (X_{\xi'}, T) : \xi' < \xi \rangle$.

The Furstenberg Structure Theorem (Furstenberg 1963) for topologically distal flows says:

Theorem 23. (Furstenberg) Let (X, d, T) be a minimal, topologically distal flow. Then (X, d, T) is a quasi-isometric extension of the trivial flow.

It is natural to ask what ordinals arise as η in the Furstenberg Structure Theorem. Given a distal flow (X, d, T) define the *distal height* of the flow as the least ordinal η such that (X, d, T) is topologically equivalent to a quasi-isometric extension of the trivial flow represented as a tower of extensions of height η .

It was shown in Beleznyay and Foreman (1995) that the collection of distal heights of minimal distal flows on compact metric spaces is exactly the collection of countable ordinals. The following theorem (Beleznyay and Foreman 1995) summarizes these results:

Theorem 24. (Beleznyay-Foreman) Let $\mathcal{M}$ be the Polish space of minimal flows on compact metric spaces. Then:

1. The collection of distal flows is a complete co-analytic set; in particular, it is non-Borel.
2. The natural function that associates to each distal (X, d, T) , its distal height is a Π_1^1 -norm.

The first item is proved by showing that there is a Borel reduction from the set of codes for well-orderings, WO to the set of distal flows as a subset of the space of minimal $\mathbb{Z}$ -flows.

If $(X_0, d_0, T_0), (X_1, d_1, T_2)$ are minimal distal flows on compact spaces, set $(X_0, d_0, T_0) \leq_D (X_1, d_1, T_2)$ if the distal height of (X_0, d_0, T_0) is less than or equal to the distal height of (X_1, d_1, T_2) . Then the theorem shows that $\leq_D$ is a true Π_1^1 -

norm on the collection of distal flows. It follows that for all countable ordinals η there is a distal flow of distal height η and that the collection of minimal distal homeomorphisms of distal height less than η is a Borel set.

In the setting of measure-preserving transformations Separate work of Furstenberg and Zimmer (Furstenberg 1981; Zimmer 1976) produced analogues of the topological Furstenberg Structure Theorem that hold for measure-preserving transformations. We describe it here without giving details and with very slight inaccuracies.

The category we are working in consists of measure-preserving transformations of a standard measure space, which we can take to be the unit interval with Lebesgue measure. By Halmos' theorem, the ergodic diffeomorphisms form a dense $\mathcal{G}_\delta$ set. Hence, the weak topology makes the ergodic transformations a Polish space. We take our ambient Polish space $EMPT$ to be the set of ergodic measure-preserving transformations of the unit interval.

Let $(Y, \mathcal{C}, \nu, S)$ be an ergodic measure-preserving transformation, and G a compact group. Let H be a compact subgroup of G . Then the quotient space G/H carries a left-invariant Haar measure. A compact co-cycle extension

$$S_\phi : Y \times G/H \rightarrow Y \times G/H$$

of Y is determined by a measurable function $\phi : Y \rightarrow G/H$ and given by the formula:

$$S_\phi(y, [g]_H) = (S(y), [\phi(y)g]_H).$$

Then S_ϕ preserves the product of ν and Haar measure on G/H .

Definition 25. (See, e.g., Furstenberg 1981). Let $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$ be ergodic measure-preserving systems. Then:

1. X is a compact extension of Y if X is measure equivalent to a compact co-cycle extension of Y .

2. Let $\pi : X \rightarrow Y$ be a measure-preserving extension. Let $X \times_Y X$ be the measure space which has domain $\{(x_1, x_2) : \pi(x_1) = \pi(x_2)\}$ and the canonical measure projecting to ν that makes the two copies of X relatively independent over $(Y, \mathcal{C}, \nu)$. Let T be the measure-preserving system taking (x_0, x_1) to $(T(x_0), T(x_1))$. Then Y canonically a factor of $X \times_Y X$ and X is a weakly mixing extension of Y if $X \times_Y X$ is ergodic.

The theorem proved independently by Furstenberg and Zimmer is:

Theorem 26. Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure-preserving system. Then there is a countable ordinal η and a system of measure-preserving transformations $\langle (X_\alpha, \mathcal{B}_\alpha, \mu_\alpha, T_\alpha) : \alpha \leq \eta + 1 \rangle$ such that

1. $(X_0, \mathcal{B}_0, \mu_0, T_0)$ is the trivial flow.
2. For each $\alpha < \eta$, $X_{\alpha+1}$ is a compact extension of X_α .
3. If α is a limit ordinal then X_α is the inverse limit of $\langle X_\beta : \beta < \alpha \rangle$.
4. $X_{\eta+1}$ is either:
 - a trivial extension of X_η (so $X_{\eta+1} \cong X_\eta$), or
 - a weakly-mixing extension of X_η .

Note that in each case, for $\alpha = 1$, the transformation X_α has to be a translation on a compact quotient G/H . In this case, G must be abelian so G/H is also an abelian group. Hence, monothetic and an example of the form due to Halmos and von Neumann.

In analogy to the situation for topologically distal transformations, we give the following definition:

Definition 27. An ergodic measure-preserving system $(X, \mathcal{B}, \mu, T)$ is measure distal if it can be written in the form of the previous theorem with $X_{\eta+1} = X_\eta$.

As in the topological case, we define the *distal height* of a measure distal transformation $(X, \mathcal{B}, \mu, T)$ to be the least η such that

$(X, \mathcal{B}, \mu, T)$ can be written satisfying 1–4 of the theorem.

The results are now analogous to the topological case (see Beleznyay and Foreman 1996):

Theorem 28. (Beleznyay-Foreman) Let $EMPT$ $([0, 1])$ be the Polish space of ergodic measure-preserving transformations of the unit interval. Then:

1. The collection of measure distal flows is a complete co-analytic set; in particular, it is non-Borel.
2. The natural function that associates to each measure distal $(X, \mathcal{B}, \mu, T)$ its distal height is a Π_1^1 -norm.

It follows from the second item that if η is a fixed countable ordinal then the collection of measure distal homeomorphisms of distal height less than η is a Borel set.

An open problem in both settings The following problems are analogous and involve *overspill*. (See Open Problem 7)

- Let X be the collection of minimal homeomorphisms on compact metric spaces and $\mathcal{D}$ be the collection of topologically distal transformations. Is there a Borel set $B \subseteq X \times X$ such that $B \cap (\mathcal{D} \times \mathcal{D})$ is the relation of topological conjugacy?
- Let $\mathcal{MD}$ be the collection of measure distal transformations. Is there a Borel set $B \subseteq EMPT \times EMPT$ such that $B \cap (\mathcal{MD} \times \mathcal{MD})$ is the relation of measure conjugacy?

In unpublished work, Foreman showed that for the collection of measure distal transformations with height bounded by a countable ordinal α , the answer to the second question is yes.

Complexity in Classification Theory

In this section, we list various benchmarks for complexity of classifications of equivalence

relations E that have roots outside of dynamical systems. The most sensitive measures of complexity are the two dimensional questions about reductions *between equivalence relations*. The general regions used to classify equivalence relations are given in Fig. 2.

A basic one-dimensional benchmark If X is a Polish space then $X \times X$ is also Polish. For any equivalence relation, $E \subseteq X \times X$, one can ask the one-parameter question:

- Is E Borel as a subset of $X \times X$?

If not, then determining whether a pair $(x_0, x_1) \in X \times X$ is in the relation E cannot be accomplished with inherently countable resources.

Many important equivalence relations arise out of group actions. If G acts on X then the equivalence relation $x \sim y$ if and only if there is a $g \in G$ such that $gx = y$ is called the orbit equivalence relation.

A *Polish group action* is an action of a Polish group G on a Polish space X that is jointly continuous in each variable. Section “ S_∞ -Actions” is concerned with actions of the group of permutations of the natural numbers, S_∞ . When we say G acts on X where G and X are both Polish, we will always be assuming that the action is a Polish group action.

Remark. When we introduce a class of equivalence relations (such as the S_∞ -actions defined below), we really mean to take $\leq_B^2$ downwards closure. So the class we label “ S_∞ -actions” means any equivalence relation reducible to an S_∞ -action.

Equivalence Relations with Countable Classes

The equivalence relations that have finite or countable classes are, perhaps, the simplest equivalence relations. There is an extensive literature on this, see, for example, Jackson et al. (2002); Kechris and Miller (2004). Equivalence relations

with countable classes are referred to as *countable equivalence relations*.

The orbit equivalence relation of any countable group action has countable classes, so there is a plethora of examples.

Let $\langle \phi_n : n \in \mathbb{N} \rangle$ be a collection of Borel bijections of a Polish space to itself. Let Γ be the group generated by $\langle \phi_n : n \in \mathbb{N} \rangle$. Then each orbit of Γ is countable and the orbit equivalence relation is Borel. Conversely, a theorem of Feldman and Moore (1977) states:

Theorem 29. (Feldman-Moore) Suppose that $E \subseteq X \times X$ is a countable Borel equivalence relation on a Polish space X . Then there is a countable group Γ and a Borel action of Γ on X such that E is the orbit equivalence relation of the action.

A $\leq_B^2$ maximal countable Borel equivalence relation Let G be a countable discrete group and $\phi : G \times X \rightarrow X$ be a Borel action on a Polish space X . Then the orbit equivalence relation is a countable Borel equivalence relation on X .

Following (Jackson et al. 2002) we let s_G be the shift action of G on the space X^G , consisting of functions $f : G \rightarrow X$. Let $E(G, X)$ be the orbit equivalence relation of this action. Dougherty et al. (1994) showed that if G is the free group on two generators, F_2 , then $E(F_2, X)$ is a maximal Borel countable equivalence relation. The following is immediate from the Dougherty-Jackson-Kechris theorem.

Theorem 30. Let E_∞ be any countable equivalence relation bi-reducible to $E(F_2, X)$ for any Polish space X . Then every countable Borel equivalence relation is reducible to E_∞ .

We show more about countable equivalence relations in section “Equivalence Relations with Countable Classes.”

The Glimm-Effros Dichotomy: E_0

We now turn to equivalence relations with uncountable classes. Because complete numerical invariants are so useful, they are the baseline

benchmark of complexity. If Y is a Polish space, we let $=_Y \subseteq Y \times Y$ be the equivalence relation of equality on Y .

If Y is a Polish space then there is a Borel injection $g : Y \rightarrow \mathbb{R}$ (See Foreman 2010). Let f be a Borel reduction of any equivalence relation $E \subseteq X \times X$ to $(Y, =_Y)$. Composing f with g we see that $g \circ f$ is a reduction of (X, E) to $(\mathbb{R}, =_{\mathbb{R}})$. Moreover $g \circ f$ is Borel if f is. Thus, we can assume that Borel reductions to any $=_Y$ can be changed to Borel reductions to equality on the real numbers.

Definition 31. Let E be an equivalence relation on a Polish space X . Then E is smooth if $E \leq_{\mathcal{B}}^2 \mathbb{R}$.

In 1990, Harrington et al. (1990), extending earlier work of Glimm (1961) and Effros (1965), showed a very general dichotomy theorem.

Definition 32. Let E_0 be the equivalence relation on the Cantor set $\{0, 1\}^{\mathbb{N}}$ defined by setting $f E_0 g$ if and only if there is an N for all $m > N$, $f(m) = g(m)$.

Proposition 33. The equivalence relation $=$ on $\{0, 1\}^{\mathbb{N}}$ is continuously reducible to E_0 .

⊢ Fix a bijection $\langle \cdot, \cdot \rangle : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$. Define $R : \{0, 1\}^{\mathbb{N}} \rightarrow \mathbb{N}$ by setting (This simple, elegant proof was taken from notes from a lecture of Su Gao. See citation in the bibliography)

$$R(f)(\langle m, n \rangle) = f(n).$$

Since R is well-defined, $f = g$ implies $R(f) = R(g)$. On the other hand suppose for all $m > N$, $R(f)(m) = R(g)(m)$. Let $n \in \mathbb{N}$ and choose an M such that $\langle M, n \rangle > N$. Then

$$\begin{aligned} f(n) &= R(f)(\langle M, n \rangle) \\ &= R(g)(\langle M, n \rangle) \\ &= g(n). \quad \dashv \end{aligned}$$

The Harrington-Kechris-Louveau dichotomy is the following theorem.

Theorem 34. (Harrington et al. 1990) Let X be a Polish space and E a Borel equivalence relation on X . Then either:

1. E is smooth or
2. $E_0 \leq_{\mathcal{B}}^2 E$.

Corollary 35. $= \leq_{\mathcal{B}}^2 E_0$ but $E_0 \not\leq_{\mathcal{B}}^2 =$. Thus E_0 is strictly more complicated than the identity equivalence relation.

This theorem suggests an orderly stair-step pattern to the benchmarks, but this is misleading, as we shall see.

We prove the easy direction of Theorem 34, because it is most useful in dynamical systems.

1. Put a measure on $\{0, 1\}$ by weighting each element $1/2$. The product measure μ on $\{0, 1\}^{\mathbb{N}}$ is often called the *coin-flipping measure*.
2. Let $G = \Sigma_{n \in \mathbb{N}} \mathbb{Z}_2$ be the direct sum of infinitely many copies of $\mathbb{Z}_2$ (with finite support).
3. The coordinatewise action of G on $(\{0, 1\}^{\mathbb{N}}, \mathcal{B}, \mu)$ preserves μ .
4. A standard 0–1 law for this action says that if $A \subseteq \{0, 1\}^{\mathbb{N}}$ is invariant under the G -action then A either has μ -measure 0 or μ -measure 1.

Theorem 36. Suppose that E is an equivalence relation on an uncountable Polish space X and $E_0 \leq_{\mathcal{B}}^2 E$. Then E is not smooth.

⊢ Toward a contradiction, assume that E is Borel reducible to the identity equivalence relation on $\mathbb{R}$ by a map $S : X \rightarrow \mathbb{R}$. Let $R : \{0, 1\}^{\mathbb{N}} \rightarrow X$ be a reduction of E_0 to E . By composing, we see that $S \circ R$ is a reduction of E_0 to the identity relation on $[0, 1]$. It suffices to show that there can be no such reduction from E_0 to the identity relation $=$ on $[0, 1]$.

Suppose that S is such a reduction. Because G preserves E_0 , for each rational number $r \in \mathbb{R}$, $L_r = \{f : S(f) < r\}$ is G invariant, and so has measure zero or one. Let

$$x = \sup\{r : L_r \text{ has measure one}\}.$$

Then $I = \{f : S(f) = x\}$ has measure one. Hence there are $f, g \in I$ such that $f E_0 g$, a contradiction. $\dashv$

$=^+$ and the Friedman-Stanley Jump Operator

Let S_∞ be the group of permutations of the natural numbers (Some authors reserve the notation S_∞ for the group of permutations of $\mathbb{N}$ that move only finitely many points and use S^∞ for the whole group of permutations. The convention in this paper is different and is becoming universal.). Then S_∞ is a $\mathcal{G}_\delta$ subset of $\mathbb{N}^\mathbb{N}$ and so carries a Polish topology. This topology is compatible with the group structure and so we view S_∞ as a Polish group.

The following operation on analytic equivalence relations was defined by Friedman and Stanley (Friedman and Stanley 1989).

Let $E \subseteq X \times X$ be an analytic equivalence relation on a Polish space. Define E^+ to be the equivalence on $X^\mathbb{N}$ given by setting $\langle [x_n]_E : n \in \mathbb{N} \rangle E^+$ -equivalent to $\langle [y_n]_E : n \in \mathbb{N} \rangle$ if and only if

for all n there is an m $x_n E y_m$
and
for all m there is an n $y_m E x_n$.

Because the collection of Borel sets (and respectively analytic sets) are closed under countable intersections and unions, the form of the definition makes it clear that if E is Borel, then E^+ is Borel (and similarly for E being analytic). If X and Y are perfect Polish spaces, then $=_X^+$ and $=_Y^+$ are the Borel bi-reducible. Moreover starting with the identity relation $=$ iterating “ $+$ ” operation any countable ordinal number α of times leads to a sequence of equivalence $\langle E_\beta : \beta < \alpha \rangle$ such that $\beta < \beta'$ implies $E_\beta \leq_B E_{\beta'}$ but $E_{\beta'} \not\leq_B E_\beta$ (Friedman and Stanley 1989). Thus this is a strict hierarchy of height ω_1 .

If E is Borel then $\{\vec{x} \in X^\mathbb{N} : [x_n]_E \neq [x_m]_E \text{ for all } n \neq m\}$ is a Borel set. Restricting the “ $+$ ”

operation” to this set at each stage also gives a proper hierarchy in $\leq_B^2$.

The following result is proved with a similar approach to remark 36. See (Gao 2022) for a simple elegant proof of the first reduction.

Theorem 37. *The following relations hold:*

$$E_0 \leq_B^2 =^+ \text{ and } =^+ \not\leq_B^2 E_0.$$

In particular, $=^+ \not\leq_B^2 =$.

S_∞ -Actions

We now discuss S_∞ -actions, where S_∞ is the Polish group of all permutations of $\mathbb{N}$ that was defined in the previous section. The S_∞ -actions are a fundamental benchmark because their actions are ubiquitous: every non-Archimedean Polish group is isomorphic to a closed subgroup of S_∞ (See Becker and Kechris 1996).

Fix an infinite perfect Polish space X and consider $P \subseteq X^\mathbb{N}$ defined as the collection of infinite sequences $\langle x_n : n \in \mathbb{N} \rangle \in X^\mathbb{N}$ such that

$$\text{for all } n \neq m, x_n \neq x_m.$$

Then P is a perfect closed subset of $X^\mathbb{N}$, S_∞ acts on P coordinatewise and for $\vec{x}, \vec{y} \in P$ we have

$$\vec{x} =^+ \vec{y} \text{ if and only if there is a } \phi \in S_\infty, \phi \vec{x} = \vec{y}.$$

Thus the equivalence relation on P induced by $=^+$ is the orbit equivalence relation of an S_∞ -action.

Remark 38. *The equivalence relation $=^+$ restricted to P is a Borel equivalence relation. Thus there are faithful S_∞ -actions inducing Borel equivalence relations.*

Remark 38 is immediate: for $\vec{x}$ and $\vec{y}$ in P , $\vec{x} =^+ \vec{y}$ if and only if for all n there is an m with $x_n = y_m$ and for all m there is an n with $y_m = x_n$.

Notation The only example of the use of the “ $+$ -operation” in this entry will be when E is the

identity equivalence relation, $=$, on the unit circle. We will restrict $=^+$ to one-to-one sequences. In an abuse of notation, we will continue to refer to this relation as $=^+$.

Classification by countable structures We now give an example to illustrate the importance of S_∞ -actions for creating invariants for equivalence relations.

Let (X, E) be an analytic equivalence relation that has complete invariants consisting of countable groups. This means that a countable group is associated to each element of X by a Borel function in such a way that xEy just in case the group associated with x is isomorphic to the group associated with y . In the language we are using here, the equivalence relation E is being reduced to the equivalence relation on pairs of countable groups given by isomorphism. The next example makes this precise.

Example. *Countable groups are determined by the characteristic function of their multiplication. Explicitly, for G an infinite countable group we can assume that its domain is $\mathbb{N}$. Define*

$$\chi_G(l, m, n) = \begin{cases} 1 & \text{if } l *_G m = n \\ 0 & \text{otherwise} \end{cases}$$

Then each $\chi_G \in \{0, 1\}^{\mathbb{N}^3}$ and $\{\chi_G : G \text{ is a countable group}\}$ is a $\mathcal{G}_\delta$ subset of $\{0, 1\}^{\mathbb{N}^3}$, and hence form a Polish space. Call it CG .

Let S_∞ act on CG by setting $(\phi \cdot \chi_G)(l, m, n) = \chi_G(\phi(l), \phi(m), \phi(n))$. If G, H are isomorphic, let $\phi : G \rightarrow H$ be the isomorphism. Then ϕ takes the multiplication table of G to the multiplication table of H . Hence $\phi \cdot \chi_G = \chi_H$. Similarly if $\phi \cdot \chi_G = \chi_H$, then $\phi : G \rightarrow H$ is an isomorphism.

This example is clearly not specific to groups—it can be adapted to any countable structures. In particular, it applies to countable graphs viewed as elements of $\{0, 1\}^{\mathbb{N}^2}$. Let E_{groups} be the equivalence relation of isomorphism of countable groups and E_{graphs} be the equivalence relation of isomorphism of countable graphs.

Theorem 39. *Let S_∞ act on an arbitrary Polish space X and F be the resulting orbit equivalence relation. Then:*

1. (Mekler 1981) $F \leq_{\mathcal{B}}^2 E_{groups}$
2. (Friedman, Stanley, see (Gao 2009)) $F \leq_{\mathcal{B}}^2 E_{graphs}$

Thus both E_{groups} and E_{graphs} are $\leq_{\mathcal{B}}^2$ -maximal among S_∞ -actions.

These examples illustrate why equivalence relations that are reducible to S_∞ -actions are referred to as those that can be *classified by countable structures*, the name for them given in (Hjorth 2000). We note that there is varying terminology for equivalence relations that are $\leq_{\mathcal{B}}^2$ bi-reducible to maximal S_∞ -actions. They are called *maximal* or *universal* S_∞ -actions and also *Borel Complete* equivalence relations.

To illustrate the large collection of equivalence relations that arise from S_∞ -actions and are $\leq_{\mathcal{B}}^2$ maximal among S_∞ -actions we cite results of Camerlo and Gao (2001).

Recall that *almost finite* commutative C^* -algebras (commutative “AF” C^* -algebras) are those that are inverse limits of finite dimensional commutative C^* algebras. They have a classification in terms of *Bratelli diagrams* (See Camerlo and Gao 2001).

Theorem 40. (Camerlo-Gao) *The isomorphism relation among commutative AF C^* -algebras is $\leq_{\mathcal{B}}^2$ -maximal among S_∞ -actions*

A maximal S_∞ -action is NOT Borel One property that makes equivalence relations such as E_{graphs} and E_{groups} very useful is that maximal S_∞ relations are not Borel. We record this here:

Fact 41. *S_∞ -actions that are $\leq_{\mathcal{B}}^2$ -maximal have complete analytic orbit equivalence relations. Hence, they are not Borel.*

In particular, *isomorphism of countable groups*, *isomorphism of countable graphs*, and *isomorphism of AFC*-algebras* are not Borel equivalence relations.

For some applications in dynamical systems the particular case of “isomorphism of countable graphs” is a particularly convenient example of an $\leq_{\mathcal{B}}^2$ -maximal S_∞ orbit equivalence relation.

A caution We do note that not all non-trivial “classifications by algebraic structures” give equivalence relations that are $\leq_{\mathcal{B}}^2$ -maximal S_∞ actions. An example, due to Friedman and Stanley (Friedman and Stanley 1989) is the equivalence relation of *isomorphism of countable abelian torsion groups*. This is a complete analytic equivalence relation that is determined by an S_∞ -action, but is not $\leq_{\mathcal{B}}^2$ -maximal among such actions.

Benchmarks to this point The equivalence relations are diagrammed in Fig. 1:

Turbulence

To establish that an equivalence relation is reducible to an S_∞ -action, one exhibits a reduction. Hjorth (2000) devised a tool for showing that an equivalence relation is *not* reducible to an S_∞ -action (The tool was discovered before there were widespread interactions between descriptive set theory and dynamical systems, so the term *turbulence* does not signify any direct connection with chaotic behavior in the sense of dynamics).

Definition 42. Let G be a Polish Group acting continuously on a Polish Space X . Then the action is *turbulent* iff:

1. Every orbit is dense.
2. Every orbit is meager.
3. For all $x, y \in X$, $V \subset X$, $U \subset G$ open, with $x \in V$, $1 \in U$, there exists $y_0 \in [y]_G$ and $\langle g_i \rangle_{i \in \mathbb{N}} \subset U$, $\langle x_i \rangle_{i \in \mathbb{N}} \subset V$ with
 - (a) $x_0 = x$,
 - (b) $x_{i+1} = g_i x_i$
 - (c) There is a subsequence i_n such that x_{i_n} converges to y_0 .

Hjorth proved the following result:

Theorem 43. Suppose that a continuous action of G on X is turbulent. Then the orbit equivalence relation cannot be reduced to an S_∞ -action.

Further results of Hjorth show that *turbulence* almost exactly captures the property of not being reducible to an S_∞ -action (See Hjorth 2000).

Remark. The property of being turbulent is generic in the sense that if $Y \subset X$ is a G_δ comeager set that is invariant under the G -action then the orbit equivalence relation induced by G on Y is also turbulent.

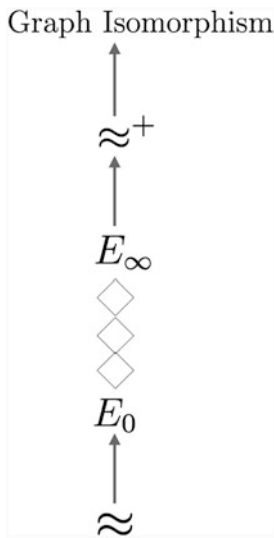
Polish Group Actions

The next major benchmark is to ask whether a given equivalence relation is reducible to the equivalence relation given by a Polish Group action. Because S_∞ is a Polish group, this is a more general collection of equivalence relations.

Because many natural equivalence relations are given by Polish group actions, this is an enormous collection. Here are some facts that help navigate the subject. An excellent book on the subject is due to Becker and Kechris (1996), although there have been many developments since the book was published.

Theorem 44. Fix a Polish group G .

1. There is a $\leq_{\mathcal{B}}^2$ maximal equivalence relation given by Polish G -actions.
2. There is a $\leq_{\mathcal{B}}^2$ maximal equivalence relation given among all Polish group actions.



The Complexity and the Structure and Classification of Dynamical Systems, Fig. 1 Benchmarks

3. If $H \subseteq G$ is a closed subgroup, then the maximal Polish H -action is continuously reducible to the maximal Polish G -action.

Items 1 and 2 are due to Becker and Kechris. Item 3 is essentially due to Mackey in the Borel case and proved by Hjorth in the continuous case.

Uspenskii proved that the group of homeomorphisms of the Hilbert cube is a universal Polish group: every Polish group is isomorphic to a closed subgroup of this group. Thus, item 2 follows by applying items 1 and 3 to the group Uspenskii used (Gao 2009; Uspenskii 1986) See Fig. 2 for a diagram of the regions discussed here.

Standard Mathematical Objects in Each Region

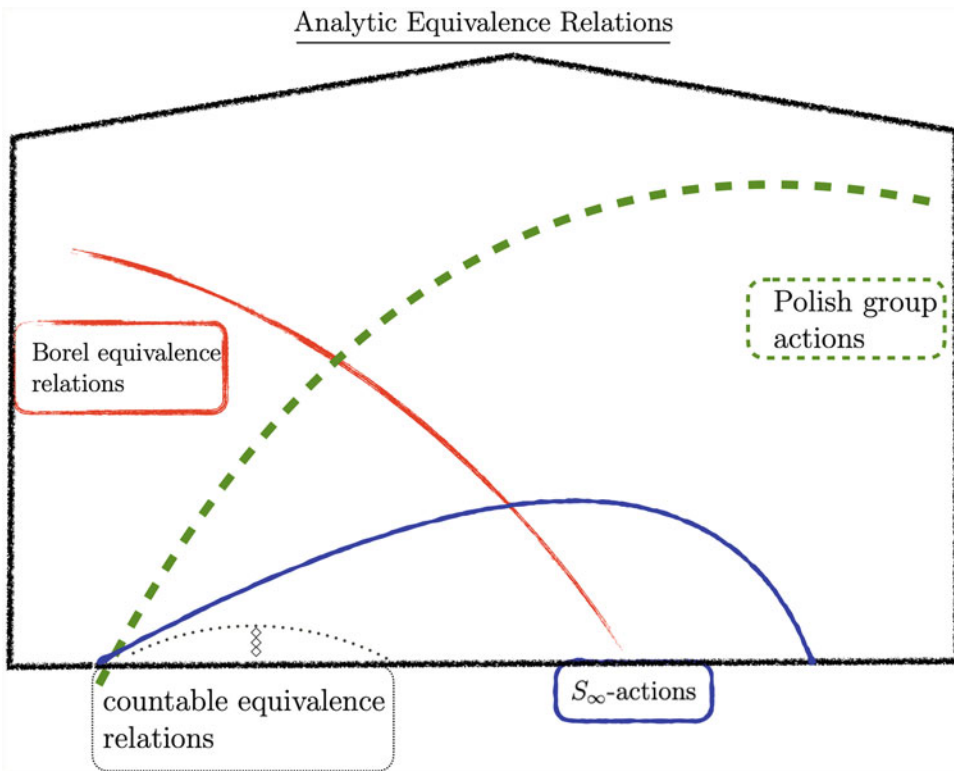
The classes with classifications in the lower regions of the diagram are extremely plentiful so

we don't expand on systems with numerical invariants or on equivalence relations with E_0 embedded in them, though we give examples of each for dynamical systems. For S_∞ we have already given an example of a Borel action (Remark 38) and $\leq_B^2$ -maximal S_∞ -actions (Theorems 39 and 40).

Countable Equivalence Relations

Recall that an equivalence relation is said to be *countable* if it has countable classes (see section "Equivalence Relations with Countable Classes")

In section "Countable Equivalence Relations", we saw that Dougherty et al. (1994) showed that there is a $\leq_B^2$ -maximal countable equivalence relation they call E_∞ . So, understanding countable equivalence relations is, in part, understanding their relationship to E_∞ . Rather than try to cover a large literature, we focus on work of Thomas (2003, 2011) that is related to dynamical systems and gives natural example of a $\leq_B^2$ -



The Complexity and the Structure and Classification of Dynamical Systems, Fig. 2 Basic regions of complexity

maximal equivalence relation with countable classes.

Note that the following sets are naturally viewed as Polish spaces endowed with countable equivalence relations:

- The space of finitely generated groups, with the equivalence relation of isomorphism denoted $\cong_{fin}$.
- The space of torsion-free abelian groups of rank n , with the equivalence relation of isomorphism denoted $\cong_n$.
- The space of torsion-free, p -local abelian groups of rank n , with the equivalence relation of isomorphism related $\cong_n^p$ (for p a prime number).

Using Zimmer's notion of *super-rigidity*, and the Kurosh-Mal'cev p -adic localization technique, Thomas (2003, 2011) proved the following results except 1 which is due to Baer. The fifth item is joint work of Thomas and Velickovic (1999) (Hjorth showed that for $n > 1$, $\cong_n \not\leq_{\mathcal{B}}^2 \cong_1$ using elementary techniques. (See Hjorth 1999.)):

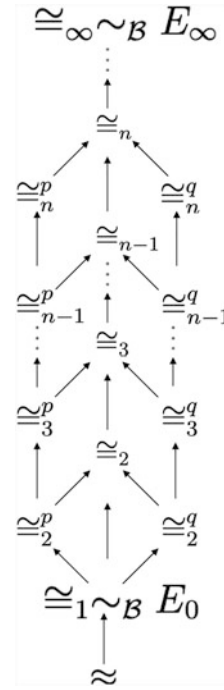
1. $\cong_1 \sim_{\mathcal{B}} E_0$,
2. For $n \geq 1$, $\cong_n$ is strictly $\leq_{\mathcal{B}}^2$ below $\cong_{n+1}$. In symbols: $\cong_n \leq_{\mathcal{B}}^2 \cong_{n+1}$,
3. For distinct primes p, q and $n \geq 2$, the relations $\cong_n^p, \cong_n^q$ are $\leq_{\mathcal{B}}^2$ -incomparable,
4. For each prime p and each $n \geq 1$, $\cong_n^p \leq_{\mathcal{B}}^2 \cong_{n+1}$ in the strict sense,
5. If $\cong_{\infty}$ is the relation of isomorphism for finitely generated groups, then $E_{\infty} \sim_{\mathcal{B}} (\cong_{\infty})$.

These results can be best summarized in Fig. 3.

General Polish Group Actions

We first give an example of a Polish group action that is not reducible to an S_{∞} -action but is still Borel.

Example. Let $\mathcal{U}$ be the collection of unitary operators on a separable Hilbert Space H and $\mathcal{N}$ be the collection of normal operators on H . Let E be the equivalence relation on $\mathcal{U}$ given by conjugacy of $\mathcal{N}$. The Spectral Theorem shows that equivalence relation of unitary conjugacy of



The Complexity and the Structure and Classification of Dynamical Systems, Fig. 3 The Thomas diagram

normal operators on a Hilbert space can be reduced to mutual absolute continuity of measures on the unit circle. The latter can be checked to be Borel using the Suslin Theorem (This was extended by Ding and Gao (2014) to the conjugacy action of all bounded linear operators on the unitary operators.).

However, it was proved by Kechris and Sofronidis that this equivalence relation is not reducible to an S_{∞} -action. They use the turbulence of the ℓ^2 -action on $\mathbb{R}^{\mathbb{N}}$. (See (Kechris and Sofronidis 2001).)

A related example is the *maximal unitary group action*. This was shown by Pestov and Gao (Gao and Pestov 2003) to lie $\leq_{\mathcal{B}}^2$ above the maximal ℓ_1 -action, which lies above all Abelian Polish group actions. In particular, it is not reducible to any ℓ_1 -action. Kanovei (2008) showed that the maximal unitary group action is not reducible to any ℓ_1 -action.

The collection of Koopman operators that arise from measure-preserving transformations on a measure space $(X, \mathcal{B}, \mu)$ is a closed subgroup of the group of unitary transformations on $L^2(X)$.

Thus, it follows from the Mackey-Hjorth theory (Theorem 44) that the isomorphism action of the measure-preserving transformations on the ergodic transformations is reducible to the maximal unitary group action.

Choquet Simplexes An example of a natural equivalence relation of maximal $\leq_B^2$ -complexity among Polish group actions is due to Sabok.

Recall that a Choquet simplex is a compact, separable, affine space X such that any point $x \in X$ is the integral of a unique measure concentrated at the extreme points of X . There is a Choquet simplex, the Poulsen simplex, which is universal in the sense that every Choquet simplex can be embedded as a face. This puts a Polish topology on the collection of Choquet simplexes. Let (C, τ) be this Polish space.

Set two Choquet simplexes S_1, S_2 equivalent if there is an affine homeomorphism taking S_1 to S_2 . Let $AFF(\text{Poulsen})$ be the group of affine homeomorphisms of the Poulsen simplex. Then $AFF(\text{Poulsen})$ is a Polish group and its natural action on C maps faces to faces, so $\text{Aff}(\text{Poulsen})$ acts on C by a Polish action.

Lindenstrauss et al. (1978) showed that any affine homeomorphism between two proper sub-faces of the Poulsen extends to an affine homeomorphism of the whole simplex. It follows that the action of $AFF(\text{Poulsen})$ on C coincides with the relation of being affinely homeomorphic. Thus, we can view the equivalence relation of being affinely homeomorphic as the orbit equivalence relation of a Polish group action.

M. Sabok (2016) showed that affine homeomorphism of Choquet simplices is $\leq_B^2$ -maximal among equivalence relations induced by Polish group actions:

Theorem 45. (Sabok) *Every orbit equivalence relation of a Polish group action is $\leq_B^2$ -reducible to the equivalence relation of affine homeomorphism of Choquet simplices.*

Borel Equivalence Relations that Don't Arise from a Polish Group Actions

Kechris and Louveau defined a Borel equivalence relation E_1 on $\{0, 1\}^{\mathbb{N} \times \mathbb{N}}$ that is not reducible to

any Polish group action and is, moreover, $\leq_B^2$ -minimal with these properties.

View $\{0, 1\}^{\mathbb{N} \times \mathbb{N}}$ as the collection of infinite sequences $\langle x_n : n \in \mathbb{N} \rangle$ where $x_n \in \{0, 1\}^{\mathbb{N}}$. Set $\langle x_n \rangle_n E_1 \langle y_m \rangle_m$ if and only if there is an N for all $n > N, x_n = y_n$. Note that this is an analogue of E_0 where elements of $\{0, 1\}$ have been replaced by elements of $\{0, 1\}^{\mathbb{N}}$ and the equivalence relation is eventual equality.

Theorem 46. (Kechris-Louveau, (Kechris and Louveau 1997)) *E_1 is Borel and not reducible to a Polish group action.*

We now give another example arising in topology. Let X be the collection of compact separable metric spaces with $\mathbb{N}$ as a dense subset. Then X carries a natural Polish topology.

We define the notion of *bi-Lipschitz equivalence*. For compact separable metric $\mathcal{K}_1, \mathcal{K}_2$ set $\mathcal{K}_1 R \mathcal{K}_2$ if and only if there is a bijection $f : \mathcal{K}_1 \rightarrow \mathcal{K}_2$ such that both f, f^{-1} are Lipschitz. The following theorem appears in Rosendal (2005).

Theorem 47. (Rosendal) *The equivalence relation $R \subseteq X \times X$ is Borel but not reducible to a Polish group action.*

The Rosendal theorem is proved by reducing E_1 to bi-Lipschitz equivalence and applying Theorem (Kechris and Louveau 1997). It is conjectured that E_1 is $\leq_B^1$ -below every Borel equivalence relation that is not reducible to a Polish group action. This is Open Problem 16.

The Maximal Analytic Equivalence Relation

While it was a folklore result that there is a maximal equivalence relation among all analytic equivalence relation, the construction was quite abstract. In Ferenczi et al. (2009), Ferenczi, Louveau, and Rosendal gave a natural example. It is an equivalence relation on the Polish space of separable Banach spaces, which can be viewed as the collection of closed subspaces of $\mathbb{C}([0, 1])$ endowed with the Effros Borel structure.

Theorem 48. (Ferenczi, Louveau, Rosendal) *The relation of isomorphism between separable*

Banach spaces is maximal among all analytic equivalence relations.

Since there are equivalence relations that are not reducible to Polish group actions, and the Ferenczi, Louveau, Rosendal example is maximal, it follows that the Ferenczi, Louveau, Rosendal example does not arise as a Polish group action.

We can now put these examples into the basic benchmark regions (Fig. 4):

Placing Dynamical Systems in Each Region

Measure Isomorphism

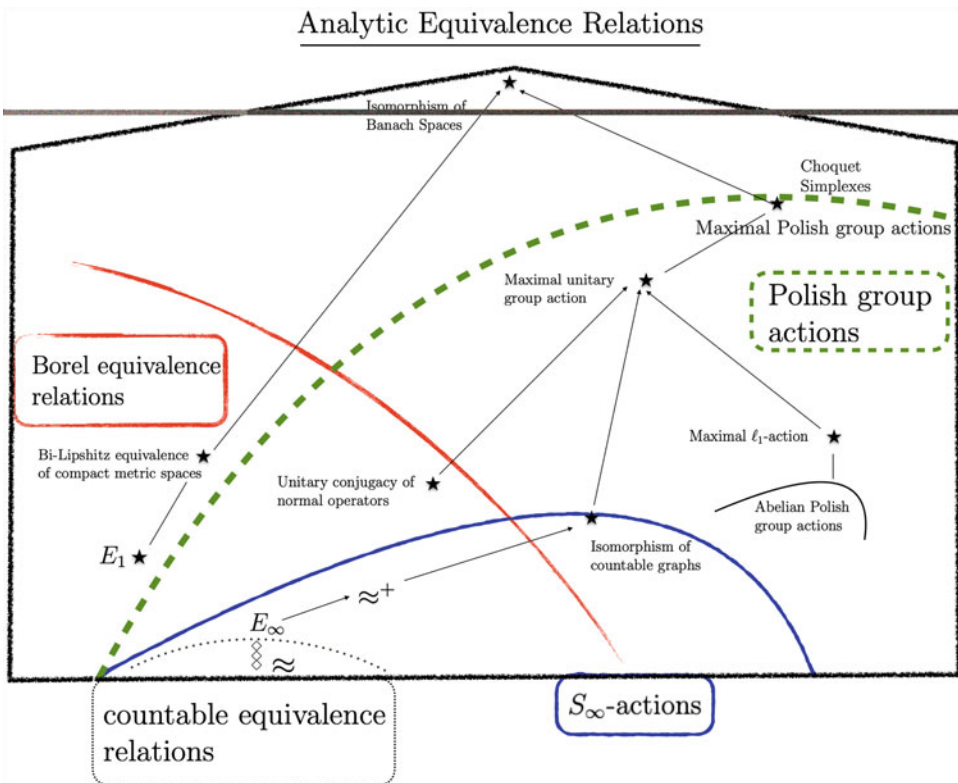
The relationship between measure isomorphism and descriptive complexity is better understood

than the relationship between topological conjugacy and descriptive set theory.

Work of Mitin (1998) and Ryzhikov (1985) studies measure-preserving transformations and show that they encode the theory of Peano Arithmetic. As a result, many questions about them are recursively undecidable. In this section, we will focus on Borel reducibility.

Systems with Numerical Invariants

The clearest example here is the class of Bernoulli Shifts. As noted in section “Rolling Dice: Bernoulli Shifts”, the collection of measure-preserving transformations isomorphic to a Bernoulli shift has a structure theory, and Ornstein’s Theorem (Theorem 9) shows that entropy is a complete numerical invariant.



The Complexity and the Structure and Classification of Dynamical Systems, Fig. 4 Classes placed in regions of complexity

Reducing E_0 to Dynamical Systems

A remarkable theorem of Feldman from a paper published in 1974 (Feldman 1974) shows the following result:

Theorem 49. *The equivalence relation E_0 is continuously reducible to isomorphism of $\mathcal{K}$ -automorphisms.*

Feldman's paper predated the language and definitions given in this entry; however, his proof is easily translated into the proof given in section "The Glimm-Effros Dichotomy: E_0 " after theorem 34.

Surprisingly, it is not known if the conjugacy relation for $\mathcal{K}$ -automorphisms is Borel (See Open Problem 4). Nor is it known if it can be reduced to an S_∞ -action.

A Class Bi-reducible to $=^+$

In theorem 12, Halmos and von Neumann showed that two discrete spectrum measurepreserving transformations are isomorphic if and only if their associated Koopman operators have the same eigenvalues. In tracing the proof (see Foreman 2000), one sees that this is a reduction of isomorphism to $=^+$ on the complex unit circle.

The reduction in the other direction also uses the Halmos-von Neuman theory. First, observe that there is a dense G_δ subset I of the complex unit circle S^1 such that any $\xi_1, \xi_2, \dots, \xi_n$ which belong to I are algebraically independent.

To each $\vec{\xi} = \langle \xi_n : n \in \mathbb{N} \rangle$ with $\xi_n \in I$, let $G_{\vec{\xi}}$ be the multiplicative subgroup of S^1 generated by $\vec{\xi}$ and consider its Pontryagin dual $\widehat{G}_{\vec{\xi}}$ consisting of all characters of the group $G_{\vec{\xi}}$. Let χ be the identity map from $G_{\vec{\xi}}$ to the unit circle. Then $\widehat{G}_{\vec{\xi}}$ is a monothetic abelian group generated by χ . Multiplication by χ in $\widehat{G}_{\vec{\xi}}$ preserves Haar measure and is an ergodic measure-preserving transformation $T_{\vec{\xi}}$. The Koopman operator of $T_{\vec{\xi}}$ has $G_{\vec{\xi}}$ as its group of eigenvalues.

Summarizing: the map that takes $\vec{\xi}$ to $(\widehat{G}_{\vec{\xi}}, \mathcal{B}, \text{Haar}, T_{\vec{\xi}})$ is a reduction of the equivalence relation $=^+$ to isomorphism of discrete

spectrum measure-preserving transformation. The upshot is the following, which appeared in Foreman (2000):

Theorem 50. *(Foreman-Louveau) Isomorphism of discrete spectrum measure-preserving transformations is Borel bi-reducible to $=^+$.*

Isomorphism Is Not Reducible to Any S_∞ -Action.

Hjorth (2001) showed that there is a turbulent equivalence relation that is $\leq_{\mathcal{B}}^2$ reducible to isomorphism for ergodic height 2 measure distal transformations (see section "Turbulence" for a discussion of turbulence). It follows that:

Theorem 51. *(Hjorth) Isomorphism for ergodic measure-preserving transformations is not reducible to an S_∞ -action.*

Foreman and Weiss (2004) improved Hjorth's result by showing that the relation of isomorphism of ergodic measure transformations is *itself* a turbulent equivalence relation.

Theorem 52. *(Foreman, Weiss) Let $MPT([0, 1])$ be the group of measure-preserving actions on $[0, 1]$ with the usual Polish topology and $EMPT([0, 1])$ be the collection of ergodic transformations. Then the conjugacy action of $MPT([0, 1])$ on $EMPT([0, 1])$ is turbulent.*

From the remark following Theorem 43, that generic classes of measure-preserving transformations, such as the ergodic or weakly mixing transformations, are also turbulent. Hence, isomorphism is not reducible to an S_∞ -action.

Corollary 53. *Let C be a dense G_δ collection of measure-preserving transformations in $[0, 1]$. Then the isomorphism relation on C is not reducible to an S_∞ -action.*

Isomorphism Is Not Borel

In Hjorth (2001), Hjorth showed that the isomorphism relation between measure-preserving transformations is not Borel. However, his proof used non-ergodic transformations in an essential way.

Foreman, Rudolph, and Weiss showed the following results (Foreman et al. 2011).

Theorem 54. (Foreman, Rudolph, Weiss) *Let G be the group of measure-preserving transformations of $[0,1]$ with the standard Polish topology and $EMPT$ be the ergodic transformations in G . The conjugacy action on G is complete analytic. In particular, it is not Borel.*

Using the method of suspensions one can lift these results to $\mathbb{R}$ -actions (Foreman et al. 2011).

Corollary 55. *The isomorphism relation between ergodic measure-preserving $\mathbb{R}$ -flows is complete analytic, and hence not Borel.*

In unpublished work, Foreman was able to establish the following:

Theorem 56. *Isomorphism of countable graphs is reducible to isomorphism for measure-preserving transformations.*

Combining these results with Corollary 53, we get:

Corollary 57. *The equivalence relation of isomorphism of measure-preserving transformations is strictly above S_∞ -actions in $\leq_{\mathcal{B}}$.*

Theorem 56 also provides a second proof of the Kechris-Sofronidis result that the maximal unitary action is strictly above the maximal S_∞ -action.

One can ask for more information in placing the relation of measure isomorphism on $EMPT$ in the $\leq_{\mathcal{B}}$ relation. It was proved by Kechris and Tucker-Drob (2013) that the relation of *unitary equivalence of normal operators* is reducible to measure isomorphism on $EMPT$. Since the unitary equivalence is Borel and isomorphism of ergodic measures is not, unitary equivalence is strictly $\leq_{\mathcal{B}}$ below isomorphism of ergodic measures.

Theorem 58. (Kechris, Tucker-Drob) *The relation of unitary equivalence of normal operators is*

$\leq_{\mathcal{B}}$ below isomorphism of ergodic measure-preserving transformations. By Theorem 54, the relation is strict.

Shortly after this entry was written, B. Weiss was able to show the following result, as yet unpublished:

Theorem. (Weiss) *Let G be a countable group. Let X be the space of homomorphisms of G to $MPT([0, 1])$ such that the action is free and ergodic. If G of the form $H \times \mathbb{Z}$, then there is a Borel reduction from the isomorphism relation on X to the isomorphism relation on $\mathbb{Z}$ -actions.*

Corollary. *If $G = \mathbb{Z}^d$ for $d \geq 1$ then there is a Borel reduction from free measure-preserving G -actions to measure-preserving $\mathbb{Z}$ -actions.*

Smooth Measure-Preserving Transformations

Much of the motivation for studying measure isomorphism is to classify the quantitative behavior of solutions to ordinary differential equations. The reductions built in Theorems 54 and 56 take their ranges in the collection of ergodic measure-preserving transformations with non-trivial odometer factors. There are no known measure-preserving transformations on compact manifolds that have odometer factors. This is a special case of the classical *Realization Problem* (Open Problem 8). A statement of the problem appears in 1975 in a paper of Ornstein (1975), and the problem is attributed to Katok by Hasselblatt in Hasselblatt (2007).

The natural question to ask is whether the measure isomorphism relation between volume preserving diffeomorphisms on a compact manifold is Borel. Does the additional structure provided by being C^∞ give enough information about measure-preserving transformations to determine isomorphism?

The answer is no:

Theorem 59. (Foreman, Weiss) *The isomorphism relation between Lebesgue measure-preserving C^∞ transformations on the 2-torus $\mathbb{T}^2$ is complete analytic. In particular, it is not Borel.*

Remark. *The techniques of Theorem 56 adapt exactly to prove the analogous result for smooth measure-preserving transformations. However, because the isomorphism relation on smooth measure-preserving transformations is not a group action, the technique of turbulence is not directly applicable. Thus it is possible that the measure isomorphism relation on C^∞ transformations on the 2-torus is $\leqslant_{\mathcal{B}}^2$ bi-reducible with the maximal S_∞ -action. This is the content of Open Problem 11.*

The proof of Theorem 59 involves three steps:

1. Identify a class of symbolic shifts (the *circular systems*) that are realizable using the Anosov-Katok ABC method.
2. Show that the class of measure-preserving transformations containing an odometer has the same factor and isomorphism structure as the class of circular systems. They are “isomorphic” categories by an isomorphism Φ .
3. Build a reduction from the space of ill-founded trees to the odometer systems; use the isomorphism Φ to transfer the range to the circular systems and then the ABC method to realize the results as diffeomorphisms.

The proofs appear in the three papers: Foreman and Weiss (2019a); Foreman and Weiss (2019b); Foreman and Weiss (2022).

We note that these results give no upper bound on the complexity of the isomorphism relation for ergodic measure-preserving transformations, other than the obvious one—that of the maximal Polish group action. Hence, we ask the question:

Is the maximal Polish group action reducible to isomorphism of ergodic measure-preserving transformation?

This is Open Problem 12.

Kakutani Equivalence

An important relation in ergodic theory is Kakutani equivalence (see Thouvenot 2002) for a nice exposition).

Let T be a measure-preserving transformation on a standard measure space $(X, \mathcal{B}, \mu)$. Let

$A \in \mathcal{B}$ be a set of positive measure and define $\mu_A : \mathcal{P}(A) \cap \mathcal{B} \rightarrow [0, 1]$ by setting

$$\mu_A(B) = \frac{\mu(B)}{\mu(A)}.$$

By Poincaré recurrence, for almost all $x \in A$, there is an $n > 0$ such that $T^n(x) \in A$. Let $n(x)$ be the least such x and $T_A(x) = T^{n(x)}(x)$. Then $T_A : A \rightarrow A$ and is a measure-preserving transformation. If T is ergodic then T_A is ergodic.

Definition 60. *Two ergodic transformations $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, \nu, S)$ are said to be Kakutani equivalent if there are positive measure sets $A \in \mathcal{B}$ and $B \in \mathcal{C}$ such that T_A is isomorphic to S_B .*

Gerber and Kunde (2021) proved the following theorem, following the general steps of the proof of Theorem 59, but with significant improvements.

Theorem 61. *The equivalence relation of Kakutani equivalence on ergodic measure-preserving diffeomorphisms of the 2-torus is complete analytic. Hence, it is not a Borel equivalence relation.*

Because Kakutani equivalence for measure-preserving diffeomorphisms is $\leqslant_{\mathcal{B}}^2$ reducible to Kakutani equivalence for general measure-preserving transformations, the following is immediate (Chronologically, the corollary preceded the theorem):

Corollary 62. *The equivalence relation of Kakutani equivalence on the space of ergodic measure-preserving transformations on $[0, 1]$ is complete analytic and hence, not Borel.*

Gerber and Kunde also proved an $\mathbb{R}$ -flow version of Corollary 62.

Definition 63. *Let $F = \{f_t : t \in \mathbb{R}\}$ be a measure-preserving $\mathbb{R}$ -flow on (X, μ) . A reparametrization of F is a jointly measurable function $\tau : X \times$*

$\mathbb{R} \rightarrow \mathbb{R}$ such that for almost every x the map $t \mapsto \tau(x, t)$ is a homeomorphism of $\mathbb{R}$ and the map $(x, t) \mapsto f_{\tau(x, t)}(x)$ is a measure-preserving flow on X .

If $G = \{g_t : t \in \mathbb{R}\}$ is another measure-preserving flow on X , then F and G are isomorphic up to a homeomorphic time change if there is a reparameterization τ of F and a measure-preserving transformation $\phi : X \rightarrow X$ such that for all t and almost every x :

$$g_t(\phi(x)) = \phi(f_{\tau(x, t)}(x)).$$

The Gerber-Kunde techniques extend to show:

Corollary 64. *The equivalence relation of isomorphism up to a homeomorphic time change for C^∞ measure-preserving flows is complete analytic and hence not Borel.*

The proofs of the Gerber-Kunde theorems do not give information about $\leq_{\mathcal{B}}^2$ reducibility to any of the benchmark analytic equivalence relations. (This is Open Problem 13.)

The Summary Diagram

Figure 5 summarizes this section. The $\mathcal{K}$ -automorphisms are not on the diagram because, other than being above E_0 , their position is not known. For the same reason, the Kakutani equivalence relation is not on the diagram.

Topological Conjugacy

In this section, we discuss the *qualitative behavior* as proposed by Smale. We do this in both the zero-dimensional and the smooth contexts. “Equivalence” in this context means *conjugate by a homeomorphism*.

After this entry was finished, in joint work with Gorodetski, the author improved the results for smooth transformations on manifolds with the equivalence relation of *topological conjugacy* to all dimensions at least one. These will appear in a future papers.

Smooth Transformations with Numerical Invariants

As discussed in section “Smooth Dynamics”, the collection of smooth transformations that is structurally stable has complete numerical invariants. These include the classes of Anosov and Morse-Smale diffeomorphisms. It is not clear how far, if at all, these classes can be extended and still have complete numerical invariants.

Diffeomorphisms of the Torus Above E_0 .

Foreman and Gorodetski showed that there are diffeomorphisms of any manifold of dimension at least 2 that are $\leq_{\mathcal{B}}^2$ above E_0 . It follows that topological conjugacy for general diffeomorphisms is not numerically classifiable.

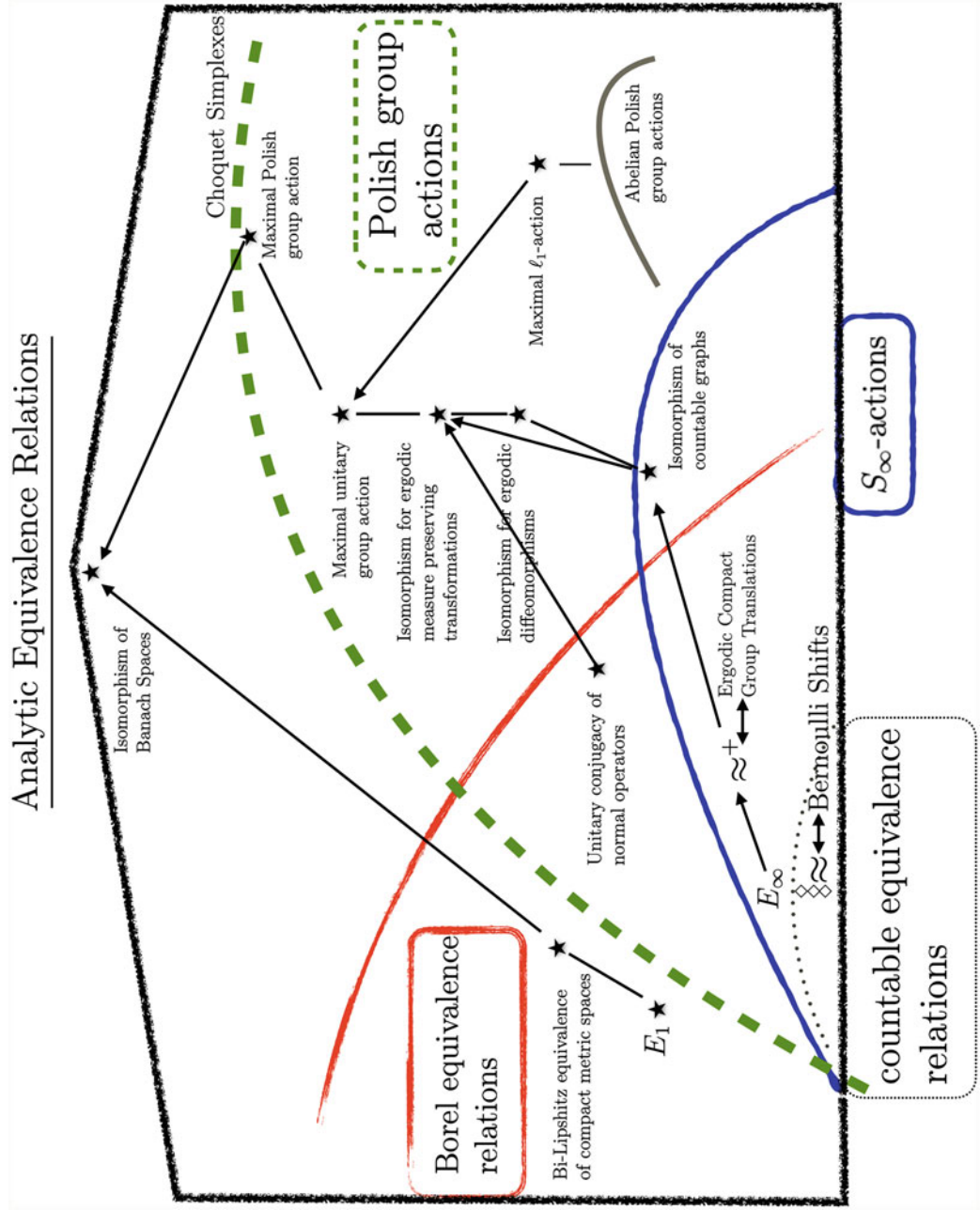
Theorem 65. (Foreman, Gorodetski) *Let M be a C^k manifold of dimension at least 2 and $1 \leq k \leq \infty$. Then, there is a continuous reduction R from E_0 to the collection of C^k diffeomorphisms with the relation of topological conjugacy.*

Remark. *In fact, the construction creates a reduction that takes values on the boundary of the Morse-Smale diffeomorphisms of the torus.*

We give here a nearly honest proof, which is relatively easy to illustrate with a picture.

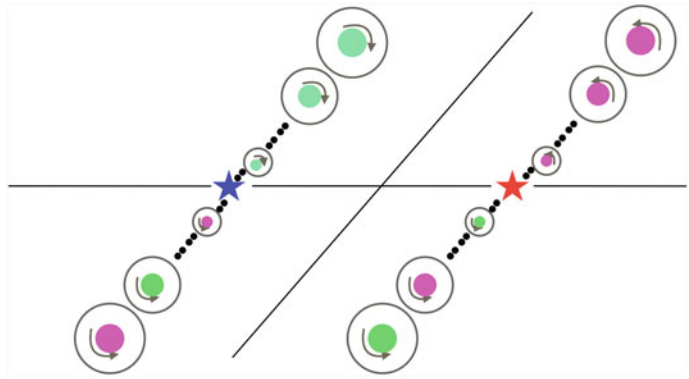
Let M be a manifold of dimension at least 2. We take it to have dimension 2 for convenience (see Fig. 6). We work inside a neighborhood diffeomorphic to $\mathbb{R}^2$. Choose two points (the left star and the right star) and four sequences of disks. The disks’ diameters go to zero at the same relatively fast rate, and two of them converge to the left point and two of them converge to the right point. There are two lines that separate the region into four parts. Each part contains exactly one of the sequences of disk. One of the lines passes through the left and right points, giving a notion of upper and lower sequences.

1. The support of the diffeomorphism is the closure of the interiors of the disks.
2. The diffeomorphism rotates each interior circle around the center of each disk and each disk contains a subdisk that the diffeomorphism



The Complexity and the Structure and Classification of Dynamical Systems, Fig. 5 Measure-preserving systems

The Complexity and the Structure and Classification of Dynamical Systems,
Fig. 6 How to embed E_0



rotates by a constant amount. (The subdisks are indicated by the shading.)

- Each of the sequence of disks above the horizontal line is identified with a fixed sequence of rotation numbers tending to zero and these are different for the two sequences.
- There is a fixed sequence of different rotation numbers with $|r_n| < 1/2$ and $\langle r_n : n \in \mathbb{N} \rangle$ tending to zero, with infinitely many positive and infinitely many negative.
- The lower subdisks have rotation numbers $\langle \pm r_n : n \in \mathbb{N} \rangle$.
- The lower subdisks then code an infinite sequence of 0's and 1's depending on whether they agree on the sign of the rotation in the n^{th} subdisk.

The reduction R maps $\{0, 1\}^{\mathbb{N}}$ to diffeomorphisms built with the two points and the sequences of disks. Given $f : \mathbb{N} \rightarrow \{0, 1\}$, $R(f)$ is the diffeomorphism that

- Rotates the n -th subdisc of lower right sequence in the positive direction and lower left sequence in the negative direction if $f(n) = 1$.
- Rotates the n^{th} subdisk of the lower left sequence in the positive direction and the n^{th} subdisk of the lower right sequence in the negative direction if $f(n) = 0$.

By construction, on the range of R , each conjugating homeomorphism of $\mathbb{R}^2$ must take the sequences of upper disks to themselves, and only

swap finitely many of the lower disks. Thus, R is a reduction of E_0 to topological conjugacy. $\dashv$

This example is, in many ways, unsatisfying. It is the identity on large portions of the space. This raises myriad questions. A simple one to frame is the following (See Open Problem 14.)

Does E_0 reduce to the collection of topologically minimal diffeomorphisms of the 2-torus with the relation of topological conjugacy?

A Dynamical Class that Is $\leq_{\mathcal{B}}^2$ Maximal for Countable Equivalence Relations

A well-studied class of dynamical systems is the collection of *finite subshifts*. These are closed, shift-invariant subsets of $\Sigma^{\mathbb{Z}}$ for a finite alphabet Σ .

Let $\mathcal{C}$ be the class of finite subshifts. If Σ is the alphabet then a *finite code* or *block code* is a function $c : \Sigma^l \rightarrow \Sigma$ for some $l \in \mathbb{N}$. Then c determines a mapping $c^* : \Sigma^{\mathbb{Z}} \rightarrow \Sigma^{\mathbb{Z}}$, where $c^*(f)$ is achieved by “starting at zero and sliding c along f in both directions.” At each location f restricted to the l -sized window is input into c which outputs a value in Σ .

If $\mathbb{K}$ and $\mathbb{K}'$ are compact subshifts and conjugate by a homeomorphism h , then that homeomorphism is given by a pair of finite codes, one for h and one for h^{-1} (See Lind and Marcus 1995). Since there are only countably many finite codes, it follows that the equivalence relation has only countable classes.

Clemens (see Clemens 2009) showed that general problem of homeomorphism is $\leq_{\mathcal{B}}^2$ –

maximal among equivalence relations with countable classes.

Theorem 66. (Clemens) *The relation of topological conjugacy between subshifts of $\Sigma^{\mathbb{Z}}$ (with Σ finite) is $\leqslant_{\mathcal{B}}^2$ maximal among analytic equivalence relations with countable classes.*

An Example of a Maximal S_{∞} -Action from Dynamical Systems

Since every perfect Polish space contains a homeomorphic copy of the Cantor set, it is natural to ask what the complexity is of the collection of homeomorphisms of the Cantor set with the relation E of topological conjugacy. This was answered by Camerlo and Gao (2001).

Theorem 67. (Camerlo-Gao) *Topological conjugacy of homeomorphisms of the Cantor set is bi-reducible with a maximal S_{∞} -action.*

As in the case of the smooth examples that exhibit high complexity (Theorems 65 and 68) the transformations in the range of the reduction are far from being minimal. Thus, we have the corresponding question: is topological conjugacy of minimal homeomorphisms of the Cantor set bi-reducible with the maximal S_{∞} -action? What about topologically transitive? (See Open Problem 15.)

Topological Conjugacy of Diffeomorphisms Is Not Borel

It might be expected that for objects as concrete as a C^k diffeomorphism of a compact manifold, topological conjugacy would be Borel. However, in dimensions 5 and above, it is known not to be the case.

Theorem 68. (Foreman, Gorodetski) *Let M be a C^k smooth manifold of dimension at least five and $1 \leq k \leq \infty$. Then the relation of Graph Isomorphism on countable graphs is reducible to topological conjugacy of C^k diffeomorphisms of M .*

Because *Graph Isomorphism* is a $\leqslant_{\mathcal{B}}^2$ -maximal S_{∞} -action, it is complete analytic. Hence, the following corollary is immediate.

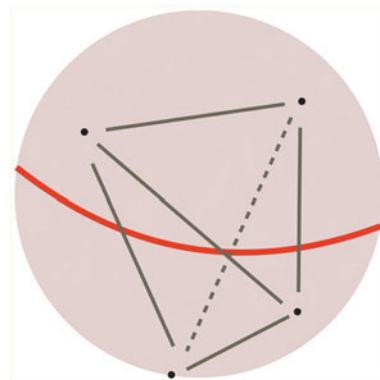
Corollary 69. *Let M be a smooth manifold of dimension at least five. Then the equivalence relation of topological conjugacy on the space of diffeomorphisms of M is complete analytic. In particular it is not Borel.*

The reduction used in the proof of Theorem 68 can be explained with pictures (although a rigorous proof requires some work).

Fix a graph G with vertices $\mathbb{N}$. Consider the solid unit ball in $\mathbb{R}^3$ and choose a set $I = \{p_n : n \in \mathbb{N}\}$ of countably many algebraically independent points tending to the South Pole. Draw line segments between each pair of points and between each point and the South Pole. The result is a complete graph on countable many points: the independent points together with the South Pole.

Let $G = (\mathbb{N}, E)$ be an arbitrary countable graph. Start the reduction by identifying $\mathbb{N}$ with the countably many points, but not the South Pole (Fig. 7).

The coding device is to put disjoint five dimensional “cigars” around each of the lines between the points in I . Given a graph G with vertices $\mathbb{N}$, the support of the diffeomorphism f_G that will be the image of G under that reduction, will be the closure of the cigars. The diffeomorphism f_G either flows toward the center of the cigar if the points connected by the cigar are connected in G or flows away from the center if they are not connected (Fig. 8).



The Complexity and the Structure and Classification of Dynamical Systems, Fig. 7 The unit ball with edges through it connecting independent points

To show this is a reduction one must show that if we are given two graphs $G = (\mathbb{N}, E)$ and $H = (\mathbb{N}, F)$, then the reduction produces diffeomorphisms f_G and f_H such that

- G is isomorphic to H if and only if
- f_G is topologically equivalent to f_H .

If f_G is topologically equivalent to f_H then the witnessing homeomorphism fixes the south pole and induces a permutation of the points in I . This, in turn, gives a permutation of the natural numbers that is an isomorphism between the graphs G and H .

The difficult direction is the opposite. Suppose that we are given a bijection $\phi : \mathbb{N} \rightarrow \mathbb{N}$ that is an isomorphism between G and H . We need to find a

homeomorphism that mimics ϕ 's behavior on the countable set I and fixes the south pole of the ball in $\mathbb{R}^3$. For a single transposition, this is easy in four dimensions (See Fig. 9). However even two transpositions can interfere with each other. Having five dimensions gives sufficient room to prevent this.

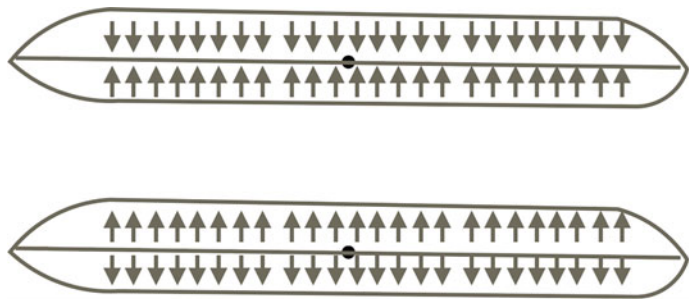
The Summary Diagram

Figure 10 summarizes the situation for topological conjugacy of diffeomorphisms, as of July 1, 2022. The section above describes what was known on February 1, 2022 and hence does not exactly correspond with the diagram.

A Descriptive Set Theory Facts

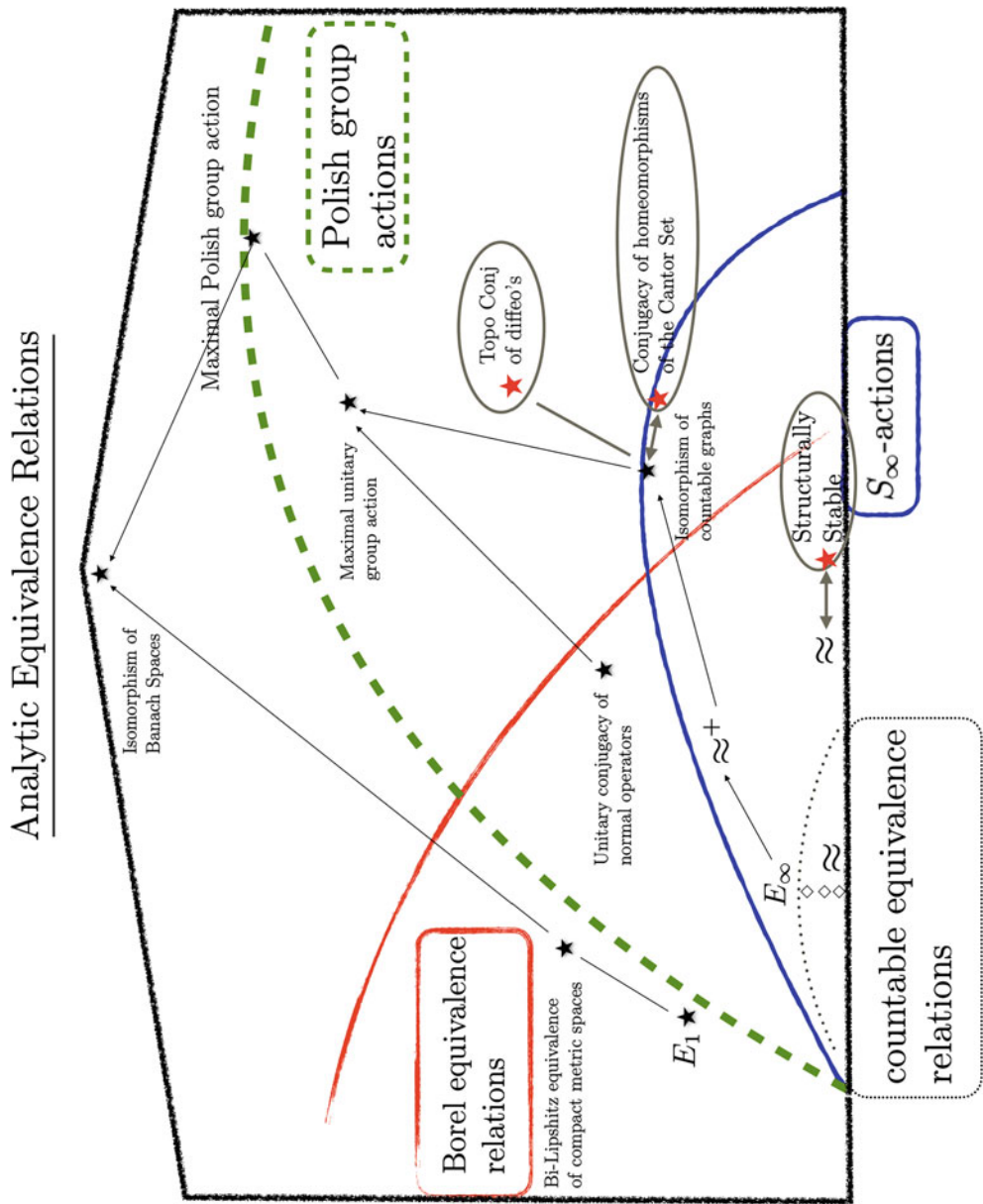
We give a very quick review of descriptive set theoretic facts used in this article. For a survey that includes complete proofs of the facts stated here,

The Complexity and the Structure and Classification of Dynamical Systems, Fig. 8 Flowing toward and away from the centerline



The Complexity and the Structure and Classification of Dynamical Systems, Fig. 9 Using the fourth dimension to swap two elements of the unit ball in $\mathbb{R}^3$





The Complexity and the Structure and Classification of Dynamical Systems, Fig. 10 Conjugacy by homeomorphisms

see the article *Naive Descriptive Set Theory* (Foreman 2010). That article does not assume any familiarity with logic.

We assume the reader is familiar with the ordinals (see Halmos' *Naive Set Theory*), which are canonical representatives for any well-ordering. Ordinality is concerned with order, as opposed to cardinality, which is concerned with size as determined by bijections. The start with the natural numbers $0, 1, 2, \dots$ which form the ordinal ω , and continue by putting a point on top to get $\omega + 1$: (Fig. 11)

After this a second copy of ω to get $\omega + \omega = \omega \times 2$. Every ordinal α has an immediate successor, $\alpha + 1$, and every collection of ordinals has a supremum. Every ordinal is either a successor ordinal or a limit ordinal. Informally, ordinals are defined to be *the set of smaller ordinals*. So, 0 is the empty set, $1 = \{0\}$ and so forth. (See (Halmos 1974; Jech 2003) or (Levy 2002) for explanation.) We will use this convention here: $\alpha = \{\beta : \beta \text{ is an ordinal and } \beta < \alpha\}$.

In ZFC, the ordinals give canonical representatives of the isomorphism types of every well-ordering. The Axiom of Choice is equivalent to the statement that every set can be well-ordered. It follows that there are ordinals of every cardinality. The smallest uncountable ordinal is called ω_1 . Because the collection of real numbers is uncountable, it has cardinality at least the cardinality of ω_1 and Cantor's *Continuum Hypothesis* says that the two cardinalities are the same.

Descriptive set theory is largely concerned with *Polish Spaces* X – these are topological spaces whose topology is compatible with a separable complete metric. A Polish space is *perfect* if it contains no isolated points. A measure on a Polish space is *standard* if open sets are measurable; it is non-atomic, complete, and separable.

Very roughly, the main topic of descriptive set theory is understanding what can be done without

the use of the uncountable Axiom of Choice. For example, the Polish spaces contain the hierarchy of *Borel Sets*, the smallest σ -algebra of subsets X containing the open sets. This algebra is built by induction on the ordinals in a hierarchy of length ω_1 . Heuristically, it contains all sets that can be built with inherently countable information.

For classifying dynamical systems, Borel sets are not sufficient. One needs *analytic* and *co-analytic* sets. These are extensions of the Borel sets that are nonetheless universally measurable.

We now give an inductive construction of the Borel sets that begins with the open sets. To keep track of the level of complexity of the set, we introduce the following “logical” notation:

Definition 70. Let (X, τ) be a Polish topological space. The levels of the Borel hierarchy are as follows:

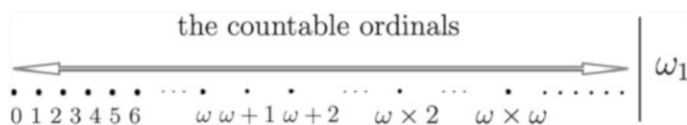
1. Σ_1^0 sets are the open subsets of X .
2. Π_1^0 sets are the closed sets.
3. Suppose the hierarchy has been defined up to (but not including) the ordinal level $\alpha < \omega_1$. Then $A \in \Sigma_\alpha^0$ if and only if there is a sequence $\langle B_i \mid i \in \omega \rangle$ of subsets of X with $B_i \in \Pi_{\beta_i}^0$ such that for each $i \in \omega$, $\beta_i < \alpha$ and

$$A = \bigcup_{i \in \omega} B_i.$$

4. $B \in \Pi_\alpha^0$ if and only if there is a subset of X , $A \in \Sigma_\alpha^0$ such that $B = X \setminus A$.

We write $\Delta_\alpha^0 = \Sigma_\alpha^0 \cap \Pi_\alpha^0$.

The second and fourth items imply that the collection defined this way is closed under complements and the third item says, in particular, that Σ_α^0 sets are constructed by taking arbitrary



The Complexity and the Structure and Classification of Dynamical Systems, Fig. 11 A few ordinals

countable unions of sets of lower complexity. Thus, Σ_α^0 sets are closed under countable unions and Π_α^0 sets are closed under countable intersections. But none of the Σ_α^0 's are closed under countable intersections, nor are the Π_α^0 's closed under unions. Thus these definitions organize the Borel sets in a hierarchy of length ω_1 . Because ω_1 has uncountable cofinality, $\bigcup_{\alpha < \omega_1} \Sigma_\alpha^0 = \bigcup_{\alpha < \omega_1} \Pi_\alpha^0$ forms a σ -algebra.

Here are some facts:

Fact. *Let X be an uncountable Polish space. Then for all countable α :*

$$\Sigma_{\alpha+1}^0 \neq \Pi_{\alpha+1}^0.$$

- Both Σ_α^0 and Π_α^0 are proper subsets of $\Delta_{\alpha+1}^0$.
- In turn, each Δ_α^0 is a proper subset of each $\Sigma_{\alpha+1}^0$ and each $\Pi_{\alpha+1}^0$.

These facts are summarized in Fig. 12.

Analytic and co-Analytic sets We now turn to the next type higher. We define two collections of universally measurable sets using projections.

Let X and Y be Polish spaces and $B \subseteq X \times Y$ be a Borel set. The set

$$A = \{x : (\text{for some } y)(x, y) \in B\} \quad (4)$$

is called an *analytic* set. It's complement

$$C = \{x : (\text{for all } y)(x, y) \notin B\} \quad (5)$$

is called a *co-analytic* set. Note that we could get an equivalent definition of coanalytic by taking a Borel set B^* and setting

$$C = \{x : (\text{for all } y)(x, y) \in B^*\}. \quad (6)$$

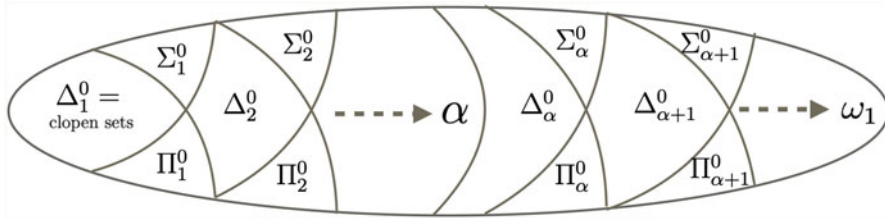
To be sure, the definition is clear – the analytic sets are all sets of the form of A in Eq. 4 with B a Borel set. The co-analytic sets are all sets of the form in Eq. 5, or equivalently all sets of the form in Eq. 6 with B a Borel set. For purposes of showing that a set is analytic or co-analytic, it is useful to use logical notation: “for some y ” can be written $\exists y$ and “for all y ” can be written $\forall y$. Informally analytic sets are “ $(\exists y)\text{Borel}$ ” and co-analytic sets are “ $(\forall y)\text{Borel}$.” Because products of Polish spaces are Polish, a set that can be written “ $(\exists y)(\exists z)\text{Borel}$ ” is analytic and “ $(\forall y)(\forall z)\text{Borel}$ ” is co-analytic.

Notation The “ Σ_1^1 sets” are the analytic sets and the “ Π_1^1 -sets” are the co-analytic sets.

Fact. *Every perfect Polish space contains a non-Borel analytic set. Moreover, the analytic sets are closed under countable intersections and unions. Hence, the coanalytic sets are also closed under unions and intersections.*

In many ways, the analytic and co-analytic sets function similarly to the open and closed sets.

Here is a remarkable theorem about the relationship between analytic and co-analytic sets due to Lusin with a very important corollary due to Suslin. (See (Lusin 1933), or section 3.5 of



The Complexity and the Structure and Classification of Dynamical Systems, Fig. 12 The construction of the Borel Sets

(Foreman 2010). Reference (Graham and Kantor 2009) gives a popular history account.)

Theorem 71. (*Lusin's Separation Theorem*) Suppose that X is a Polish space and A, B are disjoint analytic subsets of X . Then there is a Borel set C with $A \subset C$ and $C \cap B = \emptyset$. (So C separates A from B .)

The next corollary is immediate.

Corollary 72. (*Suslin's Theorem*) Let $A \subseteq X$. If both A and $X \setminus A$ are analytic, then A is Borel.

Example 73. For Borel group actions, the orbit equivalence relation is always analytic. If G is acting on X then

$$B = \{(x_0, x_1, g) : gx_0 = x_1\}$$

is a Borel set. The pair (x_0, x_1) lies in the same G -orbit if and only if $(x_0, x_1) \in A$ where

$$A = \{(x_0, x_1) : (\text{for some } g)(x_0, x_1, g) \in B\}.$$

Thus being in the same orbit is analytic.

Similarly if $\mathcal{C}$ and $\mathcal{D}$ are Polish spaces of measure-preserving transformations and $E \subseteq \mathcal{D}$ is a Borel set then

$A = \{T \in \mathcal{C} : (\text{for some measure isomorphism } \phi \text{ and some } S \in E) \phi \circ T = S \circ \phi \text{ (a.e.)}\}$ is analytic.

From this one can see that the collection of measure-preserving transformations of $[0, 1]$ that are isomorphic to a Bernoulli shift is an analytic set (details in Foreman 2000). These results are weaker than Feldman's paper where he shows that the collection of measure-preserving transformations isomorphic to a Bernoulli shift is Borel (Feldman 1974). This is discussed in section "Rolling Dice: Bernoulli Shifts."

Example 74. Let X be a compact metric space. Then the collection of distal homeomorphisms of X is a co-analytic set.

Because

$$\left\{ (T, x, y) : x \neq y \text{ and } (\text{for some } q)(\text{for all } n), d(T^n(x), T^n(y)) > 1/q \right\}.$$

is Borel,

$$\left\{ T : (\text{for all } x \neq y) \text{ for some } q \text{ for all } n, d(T^n(x), T^n(y)) > 1/q \right\}.$$

is in the form of Eq. 6. It follows that the collection of distal homeomorphisms of X is co-analytic. $\dashv$

Reductions and Hierarchies

Fact. Let X, Y be Polish spaces and $A \subseteq X, B \subseteq Y$. Let $f : X \rightarrow Y$ reduce A to B .

- If f is continuous and $B \in \Sigma_\alpha^0$ then $A \in \Sigma_\alpha^0$.
- If f is continuous and $B \in \Pi_\alpha^0$ then $A \in \Pi_\alpha^0$.
- If f is Borel and B is analytic, then A is analytic.
- If f is Borel and B is co-analytic, then A is co-analytic.

Orderings and pre-orderings Let $(I, \leq_I)$ be a linear ordering of a set of cardinality less than or equal to the cardinality of the real numbers and X a perfect Polish space. Let $C \subseteq X$ and $\pi : C \rightarrow I$ be a surjective function. Then I is coded by the set of pairs $\leq_\pi = \{(x, y) : \pi(x) \leq \pi(y)\}$ in the sense that the isomorphism type of I can be recovered from $\leq_\pi$. Note that $\leq_\pi \subseteq X \times X$, and hence can be studied as a subset of a Polish space, with no direct mention of I itself.

Of particular interest is the special case where I is an ordinal α and $\leq_I$ is the well-ordering of the ordinals less than α . In this case π itself can be recovered by $\leq_\pi$. When I is a well-ordering, the relation $\leq_\pi$ is called a *pre-well-ordering* because it is a pre-ordering and taking the quotient by the equivalence relation

$$x \sim y \text{ if and only if } (x \leq_\pi y \text{ and } y \leq_\pi x)$$

results in a well-ordering of the $\sim$ equivalence classes that is isomorphic to α . When I is a well-ordering the function π is called a *norm*.

Example. Let $WO \subseteq \{0, 1\}^{\mathbb{N} \times \mathbb{N}}$ be the set of characteristic functions of well-orderings of $\mathbb{N}$. Then WO is a complete Π_1^1 -set.

The set WO is frequently called the collection of codes for well-orderings.

For notational convenience when $\pi : C \rightarrow \alpha$ is a norm, it can be extended to all of X by adding a point, denoted ∞ to the range of π and extending π by setting $\pi(x) = \infty$ for every element of $X \setminus C$.

The following definition is difficult to digest, but the technicalities turn out to be essential.

Definition 75. Let X be a Polish space, α an ordinal, A a Π_1^1 -set and $\pi : X \rightarrow \alpha \cup \{\infty\}$ be such that $x \in A$ iff $\pi(x) < \infty$. Then π is a Π_1^1 -norm iff the relations $\leq_\pi^*$ and $<_\pi^*$ are both Π_1^1 where:

1. $x \leq_\pi^* y$ iff $x \in A$ and $(\pi(x) \leq \pi(y))$,
2. $x <_\pi^* y$ iff $x \in A$ and $(\pi(x) < \pi(y))$.

A motivated reader might ask why Π_1^1 ? The answer is that an analogous π on a Σ_1^1 set can be shown not to exist.

The next proposition explains, in part, the importance of Π_1^1 -norms.

Proposition 76. Suppose that $C \subseteq X$ is a Π_1^1 -set, with X Polish and $\pi : C \rightarrow \omega_1$ is a Π_1^1 -norm. Then $B \subseteq C$ is Borel implies that there is a countable ordinal α such that $\pi[B] \subseteq \alpha$

A corollary of this proposition is that if C is a Π_1^1 subset of a Polish space X , $\pi : X \rightarrow \omega_1 \cup \{\infty\}$ is a Π_1^1 -norm on C , and $A \subseteq X$ is a Borel set with π mapping a subset of A to a cofinal subset of ω_1 , then $A \not\subseteq C$. This phenomenon is called *overspill*.

The reduction property of co-analytic sets
The reduction property for Π_1^1 -sets is related to Π_1^1 -norms. It can be used to establish Luzin's theorem.

Theorem 77. (The reduction property for Π_1^1 -sets.) Let X be a perfect Polish space. Suppose that $A, B \subseteq X$ are Π_1^1 -sets. Then there are $A_0 \subseteq A$, $B_0 \subseteq B$ with:

$$A_0, B_0 \in \Pi_1^1$$

$$A_0 \cap B_0 = \emptyset$$

$$A_0 \cup B_0 = A \cup B.$$

Open Problems

The author has made no attempt to make a complete list of open problems – there are dozens. Nor have they attempted correct attributions. These are not even the most important problems, although some, such as Open Problem 8, are very classical and well studied. However, they ask questions raised by this survey that may not appear in other places.

Open Problem 1. Are there complete numerical invariants for orientation preserving diffeomorphisms of the circle with irrational rotation number up to conjugation by orientation preserving diffeomorphisms?

Comment: After the first draft of this entry was written, Kunde solved this negatively by showing that the equivalence relation E_0 is reducible to diffeomorphisms with Liouvillean rotation number under the equivalence relation of conjugation by diffeomorphisms.

Open Problem 2. Is there an algorithm for determining whether two subshifts of finite type are conjugate by homeomorphisms?

Open Problem 3. Is the analogue of Theorem 8 true for groups besides $\mathbb{Z}$. For example: let X be the space of Lebesgue measure-preserving actions of F_2 , the free group on two generators, on $[0, 1]$. Is the collection of actions measure isomorphic to Bernoulli F_2 -actions a Borel set?

Open Problem 4. Is measure isomorphism restricted to $\mathcal{K}$ -automorphisms a Borel equivalence relation? Can it be reduced to an S_∞ -action?

What about measure isomorphism for weakly mixing transformations? Or other classes of transformations?

Open Problem 5. Consider various classes of ergodic probability measure-preserving transformations such as weakly mixing, mixing,

0-entropy, and $\mathcal{K}$ -automorphisms. Is there a $\leq_{\mathcal{B}}^2$ relationship between any of these classes, or between one of these classes and the class of arbitrary ergodic probability measure-preserving transformations.

Open Problem 6. Is isomorphism for ergodic measure-preserving transformations on σ -finite measure spaces Borel?

What is the $\leq_{\mathcal{B}}^2$ relationship between isomorphisms of ergodic probability measure-preserving transformations and isomorphism of ergodic σ -finite measure-preserving transformations?

Open Problem 7. Do the structure theorems for distality give Borel criteria for isomorphism? More precisely:

- A. Let X be the collection of minimal homeomorphisms on compact metric spaces and $\mathcal{D}$ be the collection of topologically distal transformations. Is there a Borel set $B \subseteq X \times X$ such that $B \cap (\mathcal{D} \times \mathcal{D})$ is the relation of topological conjugacy?
- B. Let $\mathcal{MD}$ be the collection of ergodic measure distal transformations. Is there a Borel set $B \subseteq \text{EMPT} \times \text{EMPT}$ such that $B \cap (\mathcal{MD} \times \mathcal{MD})$ is the relation of measure conjugacy?

Open Problem 8. Suppose that $T: X \rightarrow X$ is an ergodic finite entropy transformation. Is there a compact manifold M with a smooth volume element v and a measure-preserving diffeomorphism S such that $(X, \mathcal{B}, \mu, T)$ is measure isomorphic to $(M, \mathcal{C}, v, S)$?

Open Problem 9. Let $\mathcal{C}$ be the collection of diffeomorphisms of a compact manifold M that are topologically conjugate to a structurally stable diffeomorphism. Let E be the relation of topological conjugacy restricted to $\mathcal{C}$. Is E Borel reducible to $=$? In other words, does E have Borel computable, complete numerical invariants?

This can also be asked for any standard class of structurally stable transformations, such as the Morse-Smale transformations. Is the conjugacy relation Borel?

Similarly in classes where there are existing invariants (such as the schemes that are used in dimension 3 in (Bonatti et al. 2019)) one can ask whether those invariants are themselves Borel.

Open Problem 10. Is the relation of topological conjugacy Borel when restricted to Axiom A diffeomorphisms of a compact surface? Can E_0 be embedded into topological conjugacy for Axiom A transformations?

Open Problem 11. Let M be a smooth manifold with a smooth volume element v . Let $\mathcal{SE}$ be the collection of smooth, ergodic, v -measure-preserving transformations of M . Is the measure isomorphism relation on $\mathcal{SE}$ bi-reducible with the maximal S_∞ -action?

Open Problem 12. (Sabok's Conjecture) Is the measure isomorphism relation on ergodic measure-preserving transformations of the unit interval $\leq_{\mathcal{B}}^2$ bi-reducible with the maximal Polish group action?

Open Problem 13. Where does the Kakutani equivalence relation on ergodic measure-preserving transformations sit among the analytic equivalence relations in $\leq_{\mathcal{B}}^2$? Is it reducible to an S_∞ -action? It is known to be a complete analytic set, hence not Borel (see Theorem 61).

In particular, what is the $\leq_{\mathcal{B}}^2$ relationship between the equivalence relation of Kakutani equivalence and isomorphism between ergodic probability measure-preserving transformations? Between Kakutani equivalence and isomorphism of ergodic measure-preserving transformations on σ -finite measure spaces?

Open Problem 14. Does E_0 reduce to the collection of topologically minimal diffeomorphisms of the 2-torus with the relation of topological conjugacy? What about topologically transitive diffeomorphisms?

Open Problem 15. Is topological conjugacy of minimal homeomorphisms of the Cantor set bi-reducible with the maximal S_∞ -action? What about topologically transitive homeomorphisms?

Open Problem 16. Suppose that F is a Borel equivalence relation on a Polish space that is not Borel reducible to the orbit equivalence relation of a Polish group action. Is $E_1 \leq_{\mathcal{B}}^2 F$?

Acknowledgments The author received help from many people in writing and editing this entry. He particularly wants to acknowledge Filippo Calderoni, Marlies Gerber, Anton Gorodetski, Philipp Kunde, Andrew Marks, Ronnie Pavlov. Alexander Kechris made invaluable corrections on an early text and correspondence with Su Gao was essential in the completion of the text. As always, Benjamin Weiss provided important suggestions and comments and provided very helpful references.

Expository Books and Articles on Analytic Equivalence Relations

1. The book by Gao (2009),
2. The books by Kechris (1995), Kechris–Miller (2004) and Becker–Kechris (1996),
3. The paper by Friedman and Stanley (1989),
4. Dave Marker’s website (Marker 2002).

The author would like to acknowledge partial support from NSF grant DMS-2100367.

Bibliography

- Aaronson J (1997) An introduction to infinite ergodic theory, Mathematical surveys and monographs, vol 50. American Mathematical Society, Providence, RI
- Barreira L, Valls C (2013) Dynamical systems, an introduction. Universitext, Springer, London
- Becker H, Kechris AS (1996) The descriptive set theory of polish group actions. London Mathematical society lecture note series. Cambridge University Press, Cambridge
- Beleznay F, Foreman M (1995) The collection of distal flows is not Borel. *Am J Math* 117(1):203–239
- Beleznay F, Foreman M (1996) The complexity of the collection of measure-distal transformations. *Ergodic Theory Dyn Syst* 16(5):929–962
- Bonatti C, Grines V, Pochinka O (2019) Topological classification of Morse–Smale diffeomorphisms on 3-manifolds. *Duke Math J* 168(13):2507–2558
- Braverman M, Yampolsky M (2009) Computability of Julia sets, Algorithms and computation in mathematics, vol 23. Springer-Verlag, Berlin
- Camerlo R, Gao S (2001) The completeness of the isomorphism relation for countable Boolean algebras. *Trans Am Math Soc* 353(2):491–518
- Chacon RV (1969) Weakly mixing transformations which are not strongly mixing. *Proc Am Math Soc* 22: 559–562
- Clemens JD (2009) Isomorphism of subshifts is a universal countable Borel equivalence relation. *Israel J. Math.* 170:113–123
- Ding L, Gao S (2014) Is there a spectral theory for all bounded linear operators? *Not Am Math Soc* 61(7): 730–735
- Dougherty R, Jackson S, Kechris AS (1994) The structure of hyperfinite Borel equivalence relations. *Trans Am Math Soc* 341(1):193–225
- Effros EG (1965) Transformation groups and C^* -algebras. *Ann Math* 81(2):38–55
- Feldman J (1974) Borel structures and invariants for measurable transformations. *Proc Am Math Soc* 46: 383–394
- Feldman J, Moore CC (1977) Ergodic equivalence relations, cohomology, and von Neumann algebras. I. *Trans Am Math Soc* 234(2):289–324
- Ferenczi V, Louveau A, Rosendal C (2009) The complexity of classifying separable Banach spaces up to isomorphism. *J Lond Math Soc* 79(2):323–345
- Foreman M (2000) A descriptive view of ergodic theory, Descriptive set theory and dynamical systems (Marseille-Luminy, 1996), London Mathematical Society lecture note series, vol 277. Cambridge Univ. Press, Cambridge, pp 87–171
- Foreman M (2010) Naive descriptive set theory, arXiv: 2110.08881, 1–53.
- Foreman M (2018) What is a Borel reduction? *Not Am Math Soc* 65(10):1263–1268
- Foreman M, Weiss B (2019a) From odometers to circular systems: a global structure theorem. *J Mod Dyn* 15: 345–423
- Foreman M, Weiss B (2019b) A symbolic representation for Anosov–Katok systems. *J Anal Math* 137(2): 603–661
- Foreman M, Weiss B (2022) Measure preserving diffeomorphisms of the torus are unclassifiable. *J Eur Math Soc (JEMS)*, 1–80.
- Foreman M, Gorodetski A (2022) Anti-classification results for smooth dynamical systems, arXiv: 2206.09322, 1–56.
- Foreman M, Weiss B (2004) An anti-classification theorem for ergodic measure preserving transformations. *J Eur Math Soc (JEMS)* 6(3):277–292
- Foreman M, Rudolph DJ, Weiss B (2011) The conjugacy problem in ergodic theory. *Ann Math* 173(3): 1529–1586
- Friedman H, Stanley L (1989) A Borel reducibility theory for classes of countable structures. *J Symb Log* 54(3): 894–914
- Furstenberg H (1963) The structure of distal flows. *Am J Math* 85:477–515
- Furstenberg H (1981) Recurrence in ergodic theory and combinatorial number theory, Porter lectures. Princeton University Press, Princeton
- Gao S (2009) Invariant descriptive set theory, A series of monographs and textbooks, vol 293. Taylor and Francis, Boca Raton
- Gao S (2022) Dynamical systems and countable structures, Lecture notes: <https://www.birs.ca/workshops/2022/22w5134/files/Su%20Gao/banff%20lecture%20gao%20nopause.pdf>
- Gao S, Pestov V (2003) On a universality property of some abelian Polish groups. *Fundam Math* 179(1):1–15
- Gerber M, Kunde P (2021) Anti-classification results for the Kakutani equivalence relation, <https://arxiv.org/abs/2109.06086v1>, 1–72.
- Glimm J (1961) Locally compact transformation groups. *TAMS* 101:124–138

- Graham L, Kantor J-M (2009) Naming infinity. Belknap Press, Cambridge
- Halmos PR (1944) In general a measure preserving transformation is mixing. *Ann Math* 45(2):786–792
- Halmos PR (1956) Lectures on ergodic theory. Chelsea Publishing Company, New York
- Halmos PR (1974) Naive set theory, Undergraduate texts in mathematics. Springer-Verlag, New York-Heidelberg
- Halmos PR and von Neumann J (1942) Operator methods in classical mechanics. II, *Ann of Math.* 43(2):332–350
- Harrington LA, Kechris AS, Louveau A (1990) A Glimm-Effros dichotomy for Borel equivalence relations. *J Am Math Soc* 3(4):903–928
- Hasselblatt B (2007) Problems in dynamical systems and related topics, Dynamics, ergodic theory, and geometry, Mathematical Sciences Research Institute Publications, vol 54. Cambridge University Press, Cambridge, pp 273–324
- Hjorth G (1999) Around nonclassifiability for countable torsion free abelian groups. In: Eklof PC, Göbel R (eds) *Abelian Groups and Modules: International conference in Dublin, August 10–14, 1998*. Birkhäuser, Basel, pp 269–292
- Hjorth G (2000) Classification and orbit equivalence relations, *Mathematical surveys and monographs*, vol 75. American Mathematical Society, Providence, RI
- Hjorth G (2001) On invariants for measure preserving transformations. *Fundam Math* 169(1):51–84
- Jackson S, Kechris AS, Louveau A (2002) Countable Borel equivalence relations. *J Math Log* 2(1):1–80
- Jech TJ (2003) Set theory: The third millenium edition. Springer, Berlin
- Kanovei V (2008) Borel equivalence relations: structure and classification, University lecture series, vol 14. American Mathematical Society, Providence, RI
- Katok A, Hasselblatt B (1995) Introduction to the modern theory of dynamical systems, *Encyclopedia of Mathematics and its applications*, vol 54. Cambridge University Press, Cambridge
- Kechris A (1995) Classical descriptive set theory, Graduate texts in mathematics, 1st edn. Springer, New York
- Kechris A (2010) Global aspects of Ergodic Group actions, *Mathematical surveys and monographs*. Am Math Soc 160
- Kechris AS, Louveau A (1997) The classification of Hypersmooth Borel equivalence relations. *J Am Mathematical Soc* 10(1):215–242
- Kechris A, Miller BD (2004) Topics in orbit equivalence, *Lecture notes in mathematics*. Springer, Berlin/Heidelberg
- Kechris AS, Sofronidis NE (2001) A strong generic ergodicity property of unitary and self-adjoint operators. *Ergodic Theory Dyn Syst* 21(5):1459–1479
- Kechris AS, Tucker-Drob RD (2013) The complexity of classification problems in ergodic theory, *Appalachian set theory 2006–2012*, London Mathematical Society lecture note series, vol 406. Cambridge University Press, Cambridge, pp 265–299
- Khinchin AI (1949) Mathematical foundations of statistical mechanics. Dover Publications, Inc., New York. Translated by G. Gamow
- King J (1986) The commutant is the weak closure of the powers, for rank-1 transformations. *Ergodic Theory Dyn Syst* 6(3):363–384
- Krieger W (1970) On entropy and generators of measure-preserving transformations. *Trans Am Math Soc* 149: 453–464
- Krieger W (1972a) Erratum to: “On entropy and generators of measure-preserving transformations”. *Trans Am Math Soc* 168:519
- Krieger W (1972b) On unique ergodicity. In: *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability*, University of California, Berkeley, 1970/1971, vol II: Probability theory, pp 327–346
- Levy A (2002) Basic set theory. Dover Publications, Inc., Mineola, Reprint of the 1979 original [Springer, Berlin; MR0533962 (80k:04001)].
- Lind D, Marcus B (1995) An introduction to symbolic dynamics and coding. Cambridge University Press, Cambridge
- Lindenstrass J, Olsen G, Sternfeld Y (1978) The Poulsen simplex. *Annales de l’institute Fourier* 28(1):91–114
- Lusin N (1933) Sur les classes des constituantes des complémentaires analytiques. *Ann Scuola Norm Super Pisa Cl Sci* 2(3):269–282
- Marker D (2002) Descriptive set theory, <http://homepages.math.uic.edu/~marker/math512/dst.pdf>, 1–56.
- Maruyama G (1949) The harmonic analysis of stationary stochastic processes. *Mem Fac Sci Kyūsyū Univ A* 4: 45–106
- Mekler AH (1981) Stability of nilpotent groups of class 2 and prime exponent. *J Symb Log* 46(4):781–788
- Mirzakhani M, Feng T (2014) Introduction to ergodic theory, https://www.mit.edu/~fengt/ergodic_theory.pdf, 1–56.
- Mitin AV (1998) Undecidability of the elementary theory of groups of measure-preserving transformations. *Mat Zametki* 63(3):414–420
- Moschovakis YN (2009) Descriptive set theory: second edition, *Mathematical surveys and monographs*, no. 155, 2nd edn. American Mathematical Society, Providence, RI
- Newhouse SE (1970) Nondensity of axiom A(a) on S^2 , global analysis. In: *Proceedings of the Symposium on Pure Mathematics*, vol XIV, Berkeley, 1968, American Mathematical Society, Providence, pp 191–202.
- Newhouse SE (1974) Diffeomorphisms with infinitely many sinks. *Topology* 13:9–18
- Ornstein D (1970a) Bernoulli shifts with the same entropy are isomorphic. *Adv Math* 4:337–352
- Ornstein D (1970b) The isomorphism problem for measure-preserving transformations. In: *Proceedings of the symposium on functional analysis*, Monterey, 1969. Academic Press, New York, pp 71–74. MR 0262463
- Ornstein D (1974) Ergodic theory, randomness, and dynamical systems, *Yale mathematical monographs*,

- No. 5. Yale University Press, New Haven/London, James K. Whittemore. Lectures in Mathematics given at Yale University
- Ornstein D (1975) Some open problems in Ergodic theory. Publications mathématiques et informatique de Rennes, no. S4
- Ornstein DS, Shields PC (1973) An uncountable family of K-automorphisms. *Adv Math* 10:63–88
- Ornstein DS, Weiss B (1974) Finitely determined implies very weak Bernoulli. *Israel J Math* 17:94–104
- Oshemkov AA, Sharko VV (1998) On the classification of Morse-Smale flows on two-dimensional manifolds. *Mat Sb* 189(8):93–140
- Peixoto MM (1973) On the classification of flows on 2-manifolds. In: *Proceedings of the Symposium dynamical systems*, University of Bahia, Salvador, 1971, pp 389–419
- Petersen KE (1983) *Ergodic theory*, Cambridge studies in advanced mathematics. Cambridge University Press, Cambridge
- Rohlin V (1948) A “general” a measure-preserving transformation is not mixing. *Doklady Akad Nauk SSSR* 60: 349–351
- Rohlin V (1967) Lectures on the entropy theory of transformations with invariant measure. *Uspehi Mat Nauk* 22(5):3–56, 137.
- Rosendal C (2005) Cofinal families of Borel equivalence relations and quasiorders. *J Symb Log* 70(4):1325–1340
- Ryzhikov VV (1985) Representation of transformations preserving the Lebesgue measure, in the form of a product of periodic transformations. *Mat. Zametki* 38(6):860–865, 957.
- Sabok M (2016) Completeness of the isomorphism problem for separable C^* -algebras. *Invent Math* 204(3): 833–868
- Smale S (1963) Dynamical systems and the topological conjugacy problem for diffeomorphisms. In: *Proceedings of the International Congress on Mathematicians*, Stockholm, 1962. Institute Mittag-Leffler, Djursholm, pp 49–496
- Smale S (1967) Differentiable dynamical systems. *Bull Am Math Soc* 73:747–817
- Soare R (1987) *Recursively enumerable sets and degrees, perspectives in mathematical logic*. Springer, Berlin/Heidelberg
- Thomas S (2003) The classification problem for torsion-free abelian groups of finite rank. *J Am Math Soc* 16(1): 233–258
- Thomas S (2011) The classification problem for S-local torsion-free abelian groups of finite rank. *Adv Math* 226(4):3699–3723
- Thomas S, Velickovic B (1999) On the complexity of the isomorphism relation for finitely generated groups. *J Algebra* 217(1):352–373
- Thouvenot J-P (2002) Entropy, isomorphism and equivalence in ergodic theory, *Handbook of dynamical systems*, vol 1A. North-Holland, Amsterdam, pp 205–238
- Uspenskiĭ VV (1986) A universal topological group with a countable basis. *Funktsional Anal i Prilozhen* 20(2): 86–87
- Walters P (1982) *An introduction to ergodic theory*, Graduate texts in mathematics. Springer, New York
- Weihrauch K (2000) *Computable analysis*, Texts in theoretical computer science. An EATCS series. Springer, Berlin, An introduction
- Wikipedia. Hamiltonian system. https://en.wikipedia.org/wiki/Hamiltonian_system
- Wikipedia. Rotation number. https://en.wikipedia.org/wiki/Rotation_number
- Zimmer RJ (1976) Ergodic actions with generalized discrete spectrum. *Ill J Math* 20(4):555–588
- Zippin L (2013) *Transformation groups*, Lectures in Topology; the University of Michigan Conference of 1940. University of Michigan Press, Ann Arbor



Ergodic Theory: Interactions with Combinatorics and Number Theory

Tom Ward
School of Mathematics, University of Leeds,
Leeds, UK

Article Outline

Glossary
Definition of the Subject
Introduction
Ergodic Theory
Frequency of Returns
Ergodic Ramsey Theory and Recurrence
Orbit-Counting as an Analogous Development
Diophantine Analysis as a Toolbox
Future Directions
Bibliography

Glossary

Almost everywhere (abbreviated a.e.) A property that makes sense for each point x in a measure space $(X, \mathcal{B}, \mu)$ is said to hold almost everywhere (or a.e.) if the set $N \subset X$ on which it does not hold satisfies $N \in \mathcal{B}$ and $\mu(N) = 0$.

Baker's theorem A lower bound for the absolute value of linear combinations of logarithms of algebraic numbers; this is a fundamental result in transcendental number theory.

Čech–Stone compactification of $\mathbb{N}$, $\beta\mathbb{N}$ A compact Hausdorff space containing $\mathbb{N}$ as a dense subset with the property that any map from $\mathbb{N}$ to a compact Hausdorff space K extends uniquely to a continuous map $\beta\mathbb{N} \rightarrow K$. This property and the fact that $\beta\mathbb{N}$ is a compact

Hausdorff space containing $\mathbb{N}$ characterizes $\beta\mathbb{N}$ up to homeomorphism.

Curvature An intrinsic measure of the curvature of a Riemannian manifold depending only on the Riemannian metric; in the case of a surface it determines whether the surface is locally convex (positive curvature), locally saddle-shaped (negative), or locally flat (zero).

Diophantine approximation Theory of the approximation of real numbers by rational numbers: how small can the distance from a given irrational real number to a rational number be made in terms of the denominator of the rational?

Diophantine problem A system of equations for which the set of integer solutions is sought. The description of the set of solutions may be qualitative (asserting that the set is finite), quantitative (bounding the number of solutions in terms of parameters like degrees), or effective (bounding the height of possible solutions, and as a result bounding both the number of solutions and their location).

Dynamical zeta function For a map with f_n points of period n , the dynamical zeta function is formally defined to be $\exp \sum_{n \geq 1} \frac{f_n}{n} z^n$ for a complex variable z . Convergence in a disc of positive radius is associated with exponential bounds on the growth in the number of periodic points.

Equidistributed A sequence is equidistributed if the asymptotic proportion of time it spends in an interval is proportional to the length of the interval.

Ergodic A measure-preserving transformation is ergodic if the only invariant functions are equal to a constant almost everywhere (a.e.); equivalently if the transformation exhibits the convergence in the quasi-ergodic hypothesis.

Ergodic theory The study of statistical properties of orbits in abstract models of dynamical systems; more generally properties of measure-preserving (semi-)group actions on measure spaces.

This entry gives a brief overview of some of the ways in which number theory and combinatorics interact with ergodic theory. The main themes are illustrated by examples related to recurrence, mixing, orbit counting, and Diophantine analysis.

Geodesic (flow) The shortest path between two points on a Riemannian manifold; such a geodesic path is uniquely determined by a starting point and the initial tangent vector to the path (that is, a point in the unit tangent bundle). The transformation on the unit tangent bundle defined by flowing along the geodesic defines the geodesic flow.

Haar measure (on a compact group) If G is a compact topological group, the unique measure μ defined on the Borel sets of G with the property that $\mu(A + g) = \mu(A)$ for all $g \in G$ and $\mu(G) = 1$.

Measure-theoretic entropy A numerical invariant of measure-preserving systems that reflects the asymptotic growth in complexity of measurable partitions refined under iteration of the map.

Mixing A measure-preserving system is mixing if measurable sets (events) become asymptotically independent as they are moved apart in time (under iteration).

Orbit Dirichlet series For a map with o_n closed orbits of length n , the Dirichlet series $\sum_{n \geq 1} \frac{o_n}{n^s}$ for a complex variable s . Convergence on a right half-plane is associated with polynomial bounds on orbit growth.

Pólya-Carlson theorem If a complex power series with integer coefficients has radius of convergence 1, then either it defines a rational function or it admits a natural boundary at its radius of convergence.

(Quasi) Ergodic hypothesis The assumption that, in a dynamical system evolving in time and preserving a natural measure, there are some reasonable conditions under which the “time average” along orbits of an observable (that is, the average value of a function defined on the phase space) will converge to the “space average” (that is, the integral of the function with respect to the preserved measure).

Recurrence Return of an orbit in a dynamical system close to its starting point infinitely often.

S-unit theorems A circle of results stating that linear equations in fields of zero characteristic have only finitely many solutions taken from finitely-generated multiplicative subgroups of the multiplicative group of the field (apart from

infinite families of solutions arising from vanishing sub-sums).

Topological entropy A numerical invariant of topological dynamical systems that measures the asymptotic growth in the complexity of orbits under iteration. The *variational principle* states that the topological entropy of a topological dynamical system is the supremum over all invariant measures of the measure-theoretic entropies of the dynamical systems viewed as measurable dynamical systems.

Definition of the Subject

Number theory is a branch of pure mathematics concerned with the properties of numbers in general, and integers in particular. The areas of most relevance to this entry are *Diophantine analysis* (the study of how real numbers may be approximated by rational numbers, and the consequences for solutions of equations in integers); *analytic number theory*, and in particular asymptotic estimates for the number of primes smaller than X as a function of X ; *equidistribution*, and questions about how the digits of real numbers are distributed. Combinatorics is concerned with identifying structures in discrete objects; of most interest here is that part of combinatorics connected with Ramsey theory, asserting that large subsets of highly structured objects must automatically contain large replicas of that structure. Ergodic theory is the study of asymptotic behavior of group actions preserving a probability measure; it has proved to be a powerful part of dynamical systems with wide applications.

Introduction

Ergodic theory, part of the mathematical study of dynamical systems, has pervasive connections with number theory and combinatorics. This entry briefly surveys how these arise through a small sample of results. Unsurprisingly, many details are suppressed, and of course the selection of topics reflects the author’s interests far more than it does the full extent of the flow of ideas

between ergodic theory and number theory. In addition, the selection of topics has been chosen in part to be complementary to those in related entries in the Encyclopedia. A particularly enormous lacuna is the theory of a class of systems called “arithmetic dynamical systems” itself; these are rational functions that map a variety to itself. The goal of the theory is to understand the arithmetic and geometry of orbits of points under iteration, and (depending on the field over which the variety is defined) it has strong connections to algebraic and arithmetic geometry. The monograph by Silverman (2007) gives a comprehensive overview.

More sophisticated aspects of this connection – in particular the connections between ergodic theory on homogeneous spaces and Diophantine analysis – are covered in the entries ► “Ergodic Theory on Homogeneous Spaces and Metric Number Theory” by Kleinbock and ► “Ergodic Theory: Rigidity” by Nițică; more sophisticated overviews of the connections with combinatorics may be found in the entry ► “Ergodic Theory: Recurrence” by Frantzikinakis and McCutcheon.

Ergodic Theory

While the early origins of ergodic theory lie in the quasi-ergodic hypothesis of classical Hamiltonian dynamics, the mathematical study of ergodic theory concerns various properties of group actions on measure spaces, including but not limited to several special branches:

1. The classical study of single measure-preserving transformations
2. Measure-preserving actions of $\mathbb{Z}^d$; more generally of countable amenable groups
3. Measure-preserving actions of $\mathbb{R}^d$ and more general amenable groups, called flows
4. Measure-preserving actions of lattices in Lie groups
5. Measure-preserving actions of Lie groups

We refer to the monograph of Einsiedler and the author (Einsiedler and Ward 2011) for an overview of the ergodic theory most relevant to

this entry. The ideas and conditions surrounding the quasi-ergodic hypothesis were eventually placed on a firm mathematical footing by developments starting in 1890. For a single measure-preserving transformation $T: X \rightarrow X$ of a probability space $(X, \mathcal{B}, \mu)$, Poincaré (1890) showed a *recurrence theorem*: if $E \in \mathcal{B}$ is any measurable set, then for a.e. $x \in E$ there is an infinite set of return times, $0 < n_1 < n_2 < \dots$ with $T^{n_j}(x) \in E$ (of course Poincaré noted this in a specific setting, concerned with a natural invariant measure for the “three-body” problem in planetary motion).

Poincaré’s qualitative result was made quantitative in the 1930s, when von Neumann (1932) used the approach of Koopman (1931) to show the *mean ergodic theorem*: if $f \in L^2(\mu)$ then there is some $\bar{f} \in L^2(\mu)$ for which

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n - \bar{f} \right\|_2 \rightarrow 0 \text{ as } N \rightarrow \infty;$$

clearly $\bar{f}$ then has the property that $\|\bar{f} - \bar{f} \circ T\|_2 = 0$ and $\int \bar{f} d\mu = \int f d\mu$. Around the same time, Birkhoff (1931) showed the more delicate *pointwise ergodic theorem*: for any $g \in L^1(\mu)$ there is some $\bar{g} \in L^1(\mu)$ for which

$$\frac{1}{N} \sum_{n=0}^{N-1} g(T^n x) \rightarrow \bar{g}(x) \text{ a.e.};$$

again it is then clear that $\bar{g}(Tx) = \bar{g}(x)$ a.e. and $\int \bar{g} d\mu = \int g d\mu$.

The map T is called *ergodic* if the invariance condition forces the function ($\bar{f}$ or $\bar{g}$) to be equal to a constant a.e. Thus an ergodic map has the property that the *time* or *ergodic average* $(1/N) \sum_{n=0}^{N-1} f \circ T^n$ converges to the *space average* $\int f d\mu$. An overview of ergodic theorems and their many extensions may be found in the article “Ergodic Theorems” by del Junco (2008).

Thus ergodic theory at its most basic level makes strong statements about the asymptotic behavior of orbits of a dynamical system as seen by *observables* (measurable functions on the space X). Applying the ergodic theorem to the

indicator function of a measurable set A shows that ergodicity guarantees that a.e. orbit spends an asymptotic proportion of time in A equal to the volume $\mu(A)$ of that set (as measured by the invariant measure). This points to the start of the pervasive connections between ergodic theory and number theory – but as this and other articles relate, the connections extend far beyond this.

Frequency of Returns

In this section, we illustrate the way in which a dynamical point of view may unify, explain, and extend quite disparate results from number theory.

Normal Numbers

Borel (1909) showed (as a consequence of what became the Borel–Cantelli Lemma in probability) that a.e. real number (with respect to Lebesgue measure) is *normal* to every base: that is, has the property that any block of k digits in the base- r expansion appears with asymptotic frequency r^{-k} .

Continued Fraction Digits

Analogues of normality results for the continued fraction expansion of real numbers were found by Khinchin, Kuz'min, Lévy, and others. Any irrational $x \in [0, 1]$ has a unique expansion as a continued fraction

$$x = \frac{1}{a_1(x) + \frac{1}{a_2(x) + \frac{1}{a_3(x) + \dots}}}$$

and, just as in the case of the familiar base- r expansion, it turns out that the digits $(a_n(x))$ obey precise statistical rules for a.e. x . Gauss conjectured that the appearance of individual digits would obey the law

$$\frac{1}{N} \left| \{k : 1 \leq k \leq N, a_k(x) = j\} \right| \rightarrow \frac{2 \log(1+j) - \log j - \log(2+j)}{\log 2}. \quad (1)$$

This was eventually proved by Kuz'min (1928) and Lévy (1929), and the probability distribution of the digits is the *Gauss–Kuz'min law*. Khinchin (1964) developed this further, showing for example that

$$\begin{aligned} \lim_{n \rightarrow \infty} (a_1(x) a_2(x) \dots a_n(x))^{1/n} \\ = \prod_{n=1}^{\infty} \left(\frac{(n+1)^2}{n(n+2)} \right)^{\log n / \log 2} \\ = 2.68545 \dots \text{ for a.e. } x. \end{aligned}$$

Lévy (1936) showed that the denominator $q_n(x)$ of the n th convergent $\frac{p_n(x)}{q_n(x)}$ (the rational obtained by truncating the continued fraction expansion of x at the n th term) grows at a specific exponential rate,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log q_n(x) = \frac{\pi^2}{12 \log 2} \text{ for a.e. } x.$$

There are many other expansions for real numbers using modified continued fractions; we refer to the monograph of Schweiger (1995) for more on this. There have also been several different attempts to define multidimensional continued fractions, some aimed at very difficult problems concerned with simultaneous Diophantine approximation. In contrast to the classical one-dimensional (or single variable) case, it is less clear that there is one canonical algorithm that emerges naturally. There is an extensive account of many of the multidimensional algorithms in another monograph of Schweiger (2000). Khanin et al. (2007), building on work of Lagarias (1994), described a simple multidimensional continued fraction algorithm which is related to dynamics on the homogeneous space $\mathrm{SL}_d(\mathbb{R})/\mathrm{SL}_d(\mathbb{Z})$. This does give best possible approximations to an irrational vector and is used to give theorems of KAM type with quasi-periodic behavior involving several different frequencies (it is well-known that the classical continued fraction expansion is intimately related to the geodesic flow on the space $\mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$; we refer to Einsiedler and Ward (2011) for the details).

First Digits

The astronomer Newcomb (1881) noted that the first digits of large collections of numerical data that are not dimensionless have a specific and nonuniform distribution:

The law of probability of the occurrence of numbers is such that all mantissæ of their logarithms are equally probable.

This is now known as Benford's Law, following his popularization and possible rediscovery of the phenomenon (Benford 1938). In both cases, this was an empirical observation eventually made rigorous by Hill (1995). Arnold (Arnol'd and Avez 1968, App. 12) pointed out the dynamics behind this phenomena in certain cases, best illustrated by the statistical behavior of the sequence 1, 2, 4, 8, 1, 3, 6, 1, ... of first digits of powers of 2. Empirically, the digit 1 appears about 30% of the time, while the digit 9 appears about 5% of the time.

Equidistribution

Weyl (1910) (and, separately, Bohl (1909) and Sierpiński (1910)) found an important instance of *equidistribution*. Writing $\{\cdot\}$ for the fractional part, a sequence (a_n) of real numbers is said to be equidistributed modulo 1 if, for any interval $[a, b] \subset [0, 1)$,

$$\frac{1}{N} \left| \{k : 1 \leq k \leq N, \{a_k\} \in [a, b]\} \right| \rightarrow (b - a) \text{ as } N \rightarrow \infty;$$

equivalently if

$$\frac{1}{N} \sum_{k=1}^N f(a_k) \rightarrow \int_0^1 f(t) dt \text{ as } N \rightarrow \infty$$

for all continuous functions f . Weyl showed that the sequence $\{n\alpha\}$ is equidistributed if and only if α is irrational. This result was refined and extended in many directions; for example, Hlawka (1964) and others found rates for the convergence in terms of the discrepancy of the sequence, Weyl (1916) proved equidistribution for $\{n^2\alpha\}$, and Vinogradov for $\{p_n\alpha\}$ where p_n is the n th prime.

The Ergodic Context

All the results of this section are manifestations of various kinds of convergence of ergodic averages. Borel's theorem on normal numbers is an immediate consequence of the fact that Lebesgue measure on $[0, 1)$ is invariant and ergodic for the map $x \mapsto bx$ modulo 1 with $b \geq 2$. The asymptotic properties of continued fraction digits are all a consequence of the fact that the *Gauss measure* defined by

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{dx}{1+x} \text{ for } A \subset [0, 1]$$

is invariant and ergodic for the Gauss map $x \mapsto \{\frac{1}{x}\}$, and the orbit of an irrational number under the Gauss map determine the digits appearing in the continued fraction expansion much as the orbit under the map $x \mapsto bx \pmod{1}$ determines the digits in the base b expansion.

The results on equidistribution and the frequency of first digits are related to ergodic averaging of a different sort. For example, writing $R_\alpha(t) = t + \alpha$ modulo 1 for the circle rotation by α , the first digit of 2^n is the digit j if and only if

$$\log_{10} j \leq R_{\log_{10}(2)}(0) < \log_{10}(j+1).$$

Thus the asymptotic frequency of appearance concerns the orbit of a *specific point*. In order to see what this means, consider a continuous map $T: X \rightarrow X$ of a compact metric space (X, d) . The space $\mathcal{M}(T)$ of Borel probability measures on the Borel σ -algebra of (X, d) is a non-empty compact convex set in the weak*-topology, each extreme point is an ergodic measure for T , and these ergodic measures are mutually singular. If $\mathcal{M}(T)$ is not a singleton and $\mu_1, \mu_2 \in \mathcal{M}(T)$ are distinct ergodic measures, then for a continuous function f with $\int_X f d\mu_1 \neq \int_X f d\mu_2$, it is clear that the ergodic averages $\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)$ must converge to $\int_X f d\mu_1$ a.e. with respect to μ_1 and to $\int_X f d\mu_2$ a.e. with respect to μ_2 . Thus the presence of many invariant measures for a continuous map means that ergodic averages along the orbits of specific points need not converge to the space average with respect to a chosen invariant measure.

In the extreme situation of *unique ergodicity* (a single invariant measure, which is necessarily an extreme point of $\mathcal{M}(T)$ and hence ergodic), the convergence of ergodic averages is much more uniform. Indeed, if T is uniquely ergodic with $\mathcal{M}(T) = \{\mu\}$ then, for any continuous function $f: X \rightarrow \mathbb{R}$,

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \rightarrow \int_X f d\mu \text{ uniformly in } x$$

(see Oxtoby (1952)). The circle rotation R_α is uniquely ergodic for irrational α , leading to the equidistribution results.

The ergodic viewpoint on equidistribution also places equidistribution results in a wider context. Weyl's result that $\{n^2\alpha\}$ (indeed, the fractional part of any polynomial with at least one irrational coefficient) is equidistributed for irrational α was given another proof by Furstenberg (1961) using the notion of unique ergodicity. These methods were then used in the study of nilsystems (translations on quotients of nilpotent Lie groups) by Auslander et al. (1963) and Parry (1969), and these nilsystems play an essential role for polynomial (and other non-conventional) ergodic averaging (see Host and Kra (2005a) and Leibman (2005) in the polynomial case; Host and Kra (2005b) in the multiple linear case). Remarkably, nilsystems are starting to play a role within combinatorics – an example is the work on the asymptotic number of 4-step arithmetic progressions in the primes by Green and Tao (2010). Pointwise ergodic theorems have also been found along sequences other than $\mathbb{N}$, notably for integer-valued polynomials and along the primes for L^2 functions by Bourgain (1988a, b). For more details, see the survey paper of del Junco (2008) on ergodic theorems.

Ergodic Ramsey Theory and Recurrence

In 1927, van der Waerden proved a conjecture attributed to Baudet: if the natural numbers are written as a disjoint union of finitely many sets,

$$\mathbb{N} = C_1 \sqcup C_2 \sqcup \dots \sqcup C_r, \quad (2)$$

then there must be one set C_j that contains arbitrarily long arithmetic progressions. That is, there is some $j \in \{1, \dots, r\}$ such that for any $k \geq 1$ there are $a \geq 1$ and $n \geq 1$ with

$$a, a + n, a + 2n, \dots, a + (k - 1)n \in C_j.$$

The original proof appears in van der Waerden's paper (van der Waerden 1927), and there is a discussion of how he found the proof in van der Waerden (1971).

Work of Furstenberg and Weiss (1978) and others placed the theorem of van der Waerden in the context of topological dynamics, giving alternative proofs. Specifically, van der Waerden's theorem is a consequence of *topological multiple recurrence*: the return of points under iteration in a topological dynamical system close to their starting point along finite sequences of times. The same approach readily gives dynamical proofs of Rado's extension (Rado 1933) of van der Waerden's theorem, and of Hindman's theorem (Hindman 1974). The theorems of Rado and Hindman introduce a new theme: given a set $A = \{n_1, n_2, \dots\}$ of natural numbers, write $FS(A)$ for the set of numbers obtained as finite sums $n_{i_1} + \dots + n_{i_j}$ with $i_1 < i_2 < \dots < i_j$. Rado showed that for any large n there is some C_s containing some $FS(A)$ for a set A of cardinality n . Hindman showed that there is some C_s containing some $FS(A)$ for an infinite set A .

In the theorem of van der Waerden, it is clear that for any reasonable notion of “proportion” or “density,” one of the sets C_j must occupy a positive proportion of $\mathbb{N}$. A set $A \subset \mathbb{N}$ is said to have *positive upper density* if there are sequences (M_i) and (N_i) with $N_i - M_i \rightarrow \infty$ as $i \rightarrow \infty$ such that

$$\lim_{i \rightarrow \infty} \frac{1}{N_i - M_i} |\{a \in A : M_i < a < N_i\}| > 0.$$

Erdős and Turán (1936) conjectured the stronger statement that any subset of $\mathbb{N}$ with positive upper density must contain arbitrary long arithmetic progressions. This statement was shown for

arithmetic progressions of length 3 by Roth (1952) in 1952, then for length 4 by Szemerédi (1969) in 1969. The general result was eventually proved by Szemerédi (1975) in 1975 in a lengthy and extremely difficult argument.

Furstenberg saw that Szemerédi's Theorem would follow from a deep extension of the Poincaré recurrence phenomena described in section “Ergodic Theory,” and proved that extension (Furstenberg 1977) (see also the survey article by Furstenberg et al. (1982)). The *multiple recurrence* result of Furstenberg says that for any measure-preserving system $(X, \mathcal{B}, \mu, T)$ and set $A \in \mathcal{B}$ with $\mu(A) > 0$, and for any $k \in \mathbb{N}$,

$$\liminf_{N \rightarrow \infty} \frac{1}{N - M + 1} \sum_{n=M}^N \mu(A \cap T^{-n}A \cap T^{-2n}A \cap \dots \cap T^{-kn}A) > 0.$$

An immediate consequence is that in the same setting there must be some $n \geq 1$ for which

$$\mu(A \cap T^{-n}A \cap T^{-2n}A \cap \dots \cap T^{-kn}A) > 0. \quad (3)$$

A general *correspondence principle*, due to Furstenberg, shows that statements in combinatorics like Szemerédi's Theorem are equivalent to statements in ergodic theory like (3).

This opened up a significant new field of *ergodic Ramsey theory*, in which methods from dynamical systems and ergodic theory are used to produce new results in infinite combinatorics. For an overview, see the entries ▶ “Ergodic Theory on Homogeneous Spaces and Metric Number Theory” by Kleinbock, ▶ “Ergodic Theory: Rigidity” by Nițică, ▶ “Ergodic Theory: Recurrence” by Frantzikinakis and McCutcheon, and the survey articles of Bergelson (1996, 2000, 2006). The field is too large to give an overview here, but a few examples will give a flavor of some of the themes.

Call a set $R \subset \mathbb{Z}$ a *set of recurrence* if, for any finite measure-preserving invertible transformation T of a finite measure space $(X, \mathcal{B}, \mu)$ and any set $A \in \mathcal{B}$ with $\mu(A) > 0$, there are infinitely many $n \in R$ for which $\mu(A \cap T^{-n}A) > 0$. Thus,

Poincaré recurrence is the statement that $\mathbb{N}$ is a set of recurrence. Furstenberg and Katznelson (1978) showed that if $T_1, \dots, T_k$ form a family of commuting measure-preserving transformations and A is a set of positive measure, then

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(T_1^{-n}A \cap \dots \cap T_k^{-n}A) > 0.$$

This remarkable multiple recurrence implies a multidimensional form of Szemerédi's theorem. Later, Gowers found a non-ergodic proof of this (Gowers 2007).

Furstenberg also gave an ergodic proof of Sárközy's theorem (Sárközy 1978): if $p \in \mathbb{Q}[t]$ is a polynomial with $p(\mathbb{Z}) \subset \mathbb{Z}$ and $p(0) = 0$, then $\{p(n)\}_{n>0}$ is a set of recurrence. This was extended to multiple polynomial recurrence by Bergelson and Leibman (1996).

Topology and Coloring Theorems

The existence of idempotent ultrafilters in the Čech–Stone compactification $\beta\mathbb{N}$ gives rise to an algebraic approach to many questions in topological dynamics (this notion has its origins in the work of Ellis (1969)). Using these methods, results like Hindman's finite sums theorem find elegant proofs, and many new results in combinatorics have been found. For example, in the partition (2), there must be one set C_j containing a triple x, y, z solving $x - y = z^2$.

A deeper application is to improve a strengthening of Kronecker's theorem. To explain this, recall that a set S is called *IP* if there is a sequence (n_i) of natural numbers (which do not need to be distinct) with the property that S contains all the terms of the sequence and all finite sums of terms of the sequence with distinct indices. A set S is called *IP** if it has non-empty intersection with every *IP* set, and a set S is called *IP**₊ if there is some $t \in \mathbb{Z}$ for which $S - t$ is *IP**. Thus being *IP** (or *IP**₊) is an extreme form of “fatness” for a set. Now let $1, \alpha_1, \dots, \alpha_k$ be numbers that are linearly independent over the rationals, and for any $d \in \mathbb{N}$ and kd non-empty intervals $I_{ij} \subset [0, 1]$ ($1 \leq i \leq d, 1 \leq j \leq k$), let

$$D = \{n \in \mathbb{N} : \{n^i \alpha_j\} \in I_{ij} \text{ for all } i, j\}.$$

Kronecker showed that if $d = 1$ then D is non-empty; Hardy and Littlewood showed that D is infinite, and Weyl showed that D has positive density. Bergelson (2003) uses these algebraic methods to improve the result by showing that D is an IP^*_+ set.

Polynomialization and IP -sets

As mentioned above, Bergelson and Leibman (1996) extended multiple recurrence to a polynomial setting. For example, let

$$\{p_{i,j} : 1 \leq i \leq k, 1 \leq j \leq t\}$$

be a collection of polynomials with rational coefficients and $p_{i,j}(\mathbb{Z}) \subset \mathbb{Z}$, $p_{i,j}(0) = 0$. Then if $\mu(A) > 0$, we have

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu \left(\bigcap_{i=1}^k \left(\prod_{j=1}^t T_j^{p_{i,j}(n)} \right)^{-1} A \right) > 0.$$

Using the Furstenberg correspondence principle, this gives a multi-dimensional polynomial Szemerédi theorem: If $P: \mathbb{Z}^r \rightarrow \mathbb{Z}^\ell$ is a polynomial mapping with the property that $P(0) = 0$, and $F \subset \mathbb{Z}^r$ is a finite configuration, then any set $S \subset \mathbb{Z}^\ell$ of positive upper Banach density contains a set of the form $u + P(nF)$ for some $u \in \mathbb{Z}^\ell$ and $n \in \mathbb{N}$.

In a different direction, motivated in part by Hindman's theorem, the multiple recurrence results generalize to IP -sets. Furstenberg and Katznelson (1985) proved a linear IP -multiple recurrence theorem in which the recurrence is guaranteed to occur along an IP -set. A combinatorial proof of this result has been found by Nagle et al. (2006). Bergelson and McCutcheon (2000) extended these results by proving a polynomial IP -multiple recurrence theorem. To formulate this, make the following definitions. Write $\mathcal{F}$ for the family of non-empty finite subsets of $\mathbb{N}$, so that a sequence indexed by $\mathcal{F}$ is an IP -set. More generally, an $\mathcal{F}$ -sequence

$(n_\alpha)_{\alpha \in \mathcal{F}}$ taking values in an abelian group is called an IP -sequence if $n_\alpha \cup \beta = n_\alpha + n_\beta$ whenever $\alpha \cap \beta = \emptyset$. An IP -ring is a set of the form $\mathcal{F}^{(1)} = \{\cup_{i \in \beta} \alpha_i : \beta \in \mathcal{F}\}$ where $\alpha_1 < \alpha_2 < \dots$ is a sequence in $\mathcal{F}$, and $\alpha < \beta$ means $a < b$ for all $a \in \alpha$, $b \in \beta$; write $\mathcal{F}^m_{<}$ for the set of m -tuples $(\alpha_1, \dots, \alpha_m)$ from $\mathcal{F}^m$ with $\alpha_i < \alpha_j$ for $i < j$. Write $PE(m, d)$ for the collection of all expressions of

the form $T(\alpha_1, \dots, \alpha_m) = \prod_{i=1}^r T_i^{p_i} \left(\left(n_{\alpha_j}^{(b)} \right)_{1 \leq b \leq k, 1 \leq j \leq m} \right)$, $(\alpha_1, \dots, \alpha_m) \in (\mathcal{F} \cup \emptyset)^m_{<}$, where each p_i is a polynomial in a $k \times m$ matrix of variables with integer coefficients and zero constant term with degree $\leq d$. Then for every $m, t \in \mathbb{N}$, there is an IP -ring $\mathcal{F}^{(1)}$, and an $a = a(A, m, t, d) > 0$, such that, for every set of polynomial expressions $\{S_0, \dots, S_t\} \subset PE(m, d)$,

$$IP\text{-}\lim_{(\alpha_1, \dots, \alpha_m) \in (\mathcal{F}^{(1)})^m_{<}} \mu \left(\bigcap_{i=0}^t S_i(\alpha_1, \dots, \alpha_m)^{-1} A \right) > a.$$

There are a large number of deep combinatorial consequences of this result, not all of which seem accessible by other means.

Sets of Primes

In a remarkable development, Szemerédi's theorem and some of the ideas behind ergodic Ramsey theory joined results of Goldston et al. (2009) in playing a part in Green and Tao's proof (Green and Tao 2008) that the set of primes contains arbitrarily long arithmetic progressions. This profound result is surveyed from an ergodic point of view in the article of Kra (2006). As with Szemerédi's theorem itself, this result has been extended to a polynomial setting by Tao and Ziegler (2008). Given integer-valued polynomials $f_1, \dots, f_k \in \mathbb{Z}[t]$ with

$$f_1(0) = \dots = f_k(0) = 0$$

and any $\varepsilon > 0$, Tao and Ziegler proved that there are infinitely many integers x, m with $1 \leq m \leq x^\varepsilon$ for which $x + f_1(m), \dots, x + f_k(m)$ are primes. We refer to a survey of Ziegler for an overview of these developments (Ziegler 2014).

Orbit-Counting as an Analogous Development

Some of the connections between number theory and ergodic theory arise through developments that are analogous but not directly related. A remarkable instance of this concerns the long history of attempts to count prime numbers laid alongside the problem of counting closed orbits in dynamical systems.

Counting Orbits and Geodesics

Consider first the fundamental arithmetic function $\pi(X) = |\{p \leq X : p \text{ is prime}\}|$. Tables of primes prepared by Felkel and Vega in the eighteenth century led Legendre to suggest that $\pi(X)$ is approximately $x/(\log(X) - 1.08)$. Gauss, using both computational evidence and a heuristic argument, suggested that $\pi(X)$ is approximated by

$$\text{li}(X) = \int_2^X \frac{dt}{\log t}.$$

Both of these suggestions imply the well-known asymptotic formula

$$\pi(X) \sim \frac{X}{\log X}. \quad (4)$$

Riemann brought the analytic ideas of Dirichlet and Chebyshev (who used the zeta function to find a weaker version of (4) with upper and lower bounds for the quantity $\frac{\pi(X) \log(X)}{X}$) to bear by proposing that the zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p (1 - p^{-s})^{-1}, \quad (5)$$

already studied by Euler, would connect properties of the primes to analytic methods. An essential step in these developments, due to Riemann, is the meromorphic extension of ζ from the region $\Re(s) > 1$ in (5) to the whole complex plane and a functional equation relating the value of the extension at s to the value at $1 - s$. Moreover, Riemann showed that the extension has readily understood real zeros, and that all the other zeros he could find

were symmetric about $\Re(s) = \frac{1}{2}$. The *Riemann hypothesis* asserts that zeros in the region $0 < \Re(s) < 1$ all lie on the line $\Re(s) = \frac{1}{2}$, and this remains open.

Analytic properties of the Riemann zeta function were used by Hadamard and de la Vallée Poussin to prove (4), the *Prime Number Theorem*, in 1896. Tauberian methods developed by Wiener and Ikehara (Wiener 1932) later gave different approaches to the Prime Number Theorem.

These ideas initiated the widespread use of zeta functions in several parts of mathematics, but it was not until the middle of the twentieth century that Selberg (1956) introduced a zeta function dealing directly with quantities arising in dynamical systems: the lengths of closed geodesics on surfaces of constant curvature -1 . The geodesic flow acts on the unit tangent bundle to the surface by moving a point and unit tangent vector at that point along the unique geodesic they define at unit speed. Closed geodesics are then in one-to-one correspondence with periodic orbits of the associated geodesic flow on the unit tangent bundle, and it is in this sense that the quantities are dynamical. The function defined by Selberg takes the form

$$Z(s) = \prod_{\tau} \prod_{k=0}^{\infty} \left(1 - e^{-(s+k)|\tau|}\right),$$

in which τ runs over all the closed geodesics, and $|\tau|$ denotes the length of the geodesic. In a direct echo of the Riemann zeta function, Selberg found an analytic continuation to the complex plane and showed that the zeros of Z lie on the real axis or on the line $\Re(s) = \frac{1}{2}$ (the analogue of the Riemann hypothesis for Z ; see also the paper of Hejhal (1976)). The zeros of Z are closely connected to the eigenvalues for the Laplace–Beltrami operator, and thus give information about the lengths of closed geodesics via Selberg’s trace formula in the same paper. Huber (1959) and others used this approach to give an analogue of the prime number theorem for closed geodesics – a *prime orbit theorem*.

Sinaï (1966) considered closed geodesics on a manifold M with negative curvature bounded between $-R^2$ and $-r^2$, and found the bounds

$$\begin{aligned}
 (\dim(M) - 1)r &\leq \liminf_{T \rightarrow \infty} \frac{\log \pi(T)}{T} \\
 &\leq \limsup_{T \rightarrow \infty} \frac{\log \pi(T)}{T} \leq (\dim(M) - 1)R
 \end{aligned}$$

for the number $\pi(T)$ of closed geodesics of multiplicity one with length less than T , analogous to Chebyshev's result.

The essential dynamical feature behind the geodesic flow on a manifold of negative curvature is that it is an example of an *Anosov flow* (Anosov 1967). These are smooth dynamical $\mathbb{R}$ -actions (equivalently, first-order differential equations on Riemannian manifolds) with the property that the tangent bundle has a continuously varying splitting into a direct sum $E^u \oplus E^s \oplus E^o$ and the action of the differential of the flow acts on E^u as an exponential expansion, on E^s as an exponential contraction, E^o is the one-dimensional bundle of vectors that are tangent to orbits, and the expansion and contraction factors are bounded. In the setting of Anosov flows, the natural orbit counting function is $\pi(X) = |\{\tau : \tau \text{ a closed orbit of length } |\tau| \leq X\}|$. Margulis (1969, 2004) generalized the picture to weak-mixing Anosov flows by showing a prime orbit theorem of the form

$$\pi(X) \sim \frac{e^{h_{\text{top}}X}}{h_{\text{top}}X} \quad (6)$$

for the counting function $\pi(X) = |\{\tau : \tau \text{ a closed orbit of length } |\tau| \leq X\}|$ where as before h_{top} denotes the topological entropy of the flow. Integral to Margulis' work is a result on the spatial distribution of the closed geodesics reflected in a flow-invariant probability measure, now called the Margulis measure.

Anosov also studied discrete dynamical systems with similar properties: diffeomorphisms of compact manifolds with a similar splitting of the tangent space (though in this setting E^o disappears). The archetypal Anosov diffeomorphism is a hyperbolic toral automorphism of the sort considered in section “[Orbit Growth and Convergence](#)”; for such automorphisms of the 2-torus Adler and Weiss (Adler and Weiss 1970)

constructed Markov partitions, allowing the dynamics of the toral automorphism to be modeled by a topological Markov shift, and used this to determine when two such automorphisms are measurably isomorphic. Sinai (1968), Ratner (1973), Bowen (1970, 1973), and others developed the construction of Markov partitions in general for Anosov diffeomorphisms and flows.

Around the same time, Smale (1967) introduced a more permissive hyperbolicity axiom for diffeomorphisms, Axiom A. Maps satisfying Axiom A are diffeomorphisms satisfying the same hypothesis as that of Anosov diffeomorphisms, but only on the set of points that return arbitrarily close under the action of the flow (or iteration of the map).

Thus Markov partitions, and with them associated transfer operators, became a substitute for the geometrical Laplace–Beltrami operators of the setting considered by Selberg. Bowen (1972) extended the uniform distribution result of Margulis to this setting and found an analogue of Chebyshev's theorem for closed orbits. Parry (1983) (in a restricted case) and Parry and Pollicott (1983) went on to prove the prime orbit theorem in this more general setting. The methods are an adaptation of the Ikehara–Wiener Tauberian approach to the prime number theorem.

Thus many facets of the prime number theorem story find their echoes in the study of closed orbits for hyperbolic flows: the role played by meromorphic extensions of suitable zeta functions, Tauberian methods, and so on. Moreover, related results from number theory have analogues in dynamics, for example, Mertens' theorem (Mertens 1874) in the work of Sharp (1991) and Noorani (1999) and Dirichlet's theorem in work of Parry (1984).

The “elementary” proof (not using analytic methods) of the prime number theorem by Erdős (1949) and Selberg (1949) (see the survey by Goldfeld (2004) for the background to the results, and the unfortunate priority dispute) has an echo in some approaches to orbit-counting problems from an elementary (non-Tauberian) perspective, including work of Lalley (1988) on special flows and Everest et al. (2007) in the algebraic setting.

In a different direction, Lalley (1987) found orbit asymptotics for closed orbits satisfying constraints in the Axiom A setting without using Tauberian theorems. His more direct approach is still analytic, using complex transfer operators (the same objects used by Parry and Pollicott to study the dynamical zeta function at complex values) and indeed somewhat parallels a Tauberian argument.

Further resonances with number theory arise here. For example, there are results on the distribution of closed orbits for group extensions (analogous to Chebotarev's theorem) and for orbits with homological constraints (see Sharp (1993), Katsuda and Sunada (1990)).

Of course the great diversity of dynamical systems subsumed in the phrase "prime orbit theorem" creates new problems and challenges, and in particular if there is not much geometry to work with then the reliance on Markov partitions and transfer operators makes it difficult to find higher-order asymptotics.

Dolgopyat (1998) has nonetheless managed to push the Markov methods to obtain uniform bounds on iterates of the associated transfer operators to the region $\Re(s) > \sigma_0$ with $\sigma_0 < 1$. This result has wide implications; an example most relevant to the analogy with number theory is the work of Pollicott and Sharp (1998) in which Dolgopyat's result is used to show that for certain geodesic flows there is a two-term prime orbit theorem of the form

$$\pi(X) = \text{li}(e^{h_{\text{top}}X}) + O(e^{cX})$$

for some $c < h_{\text{top}}$.

For non-positive curvature manifolds less is known: Knieper (1997) finds upper and lower bounds for the function counting closed geodesics on rank-1 manifolds of non-positive curvature of the form

$$A \frac{e^{hX}}{X} \leq \pi(X) \leq B e^{hX}$$

for constants $A, B > 0$.

Counting Orbits for Group Endomorphisms

A prism through which to view some of the deeper issues that arise in section "Counting Orbits and Geodesics" is provided by group endomorphisms. The price paid for having simple closed formulas for all the quantities involved is of course a severe loss of generality, but the diversity of examples illustrates many of the phenomena that may be expected in more general settings when hyperbolicity is lost.

Consider an endomorphism $T : X \rightarrow X$ of a compact group with the property that

$$F_n(T) = |\{x \in X : T^n x = x\}| < \infty$$

for all $n \geq 1$. The number of closed orbits of length n under T is then

$$O_n(T) = \frac{1}{n} \sum_{d|n} \mu(n/d) F_d(T). \quad (7)$$

In simple situations (hyperbolic toral automorphisms for example), it is straightforward to show that

$$\begin{aligned} \pi_T(X) &= |\{\tau : \tau \text{ a closed orbit under } T \text{ of length } \leq X\}| \\ &\sim \frac{e^{(X+1)h_{\text{top}}(T)}}{X}. \end{aligned} \quad (8)$$

Waddington (1991) considered quasi-hyperbolic toral automorphisms, showing that the asymptotic (8) in this case is multiplied by an explicit almost-periodic function bounded away from zero and infinity.

This result has been extended further into non-hyperbolic territory, which is most easily seen via the so-called connected S -integer dynamical systems introduced by Chothi et al. (1997). Fix an algebraic number field $\mathbb{K}$ with set of places $P(\mathbb{K})$ and set of infinite places $P_\infty(\mathbb{K})$, an element of infinite multiplicative order $\xi \in \mathbb{K}^*$, and a finite set $S \subset P(\mathbb{K}) \setminus P_\infty(\mathbb{K})$ with the property that $|\xi|_w \leq 1$ for all $w \notin S \cup P_\infty(\mathbb{K})$. The associated ring of S -integers is

$$R_S = \{x \in \mathbb{K} : |x|_w \leq 1 \text{ for all } w \notin S \cup P_\infty(\mathbb{K})\}.$$

Let X be the compact character group of R_S , and define the endomorphism $T: X \rightarrow X$ to be the dual of the map $x \mapsto \xi x$ on R_S . Following Weil (1967), write $\mathbb{K}_w$ for the completion at w , and for w finite, write r_w for the maximal compact subring of $\mathbb{K}_w$. Notice that if $S = P$ then $R_S = \mathbb{K}$ and $F_n(T) = 1$ for all $n \geq 1$ by the product formula for $\mathbb{A}$ -fields. As the set S shrinks, more and more periodic orbits come into being, and if S is as small as possible (given ξ) then the resulting system is (more or less) hyperbolic or quasi-hyperbolic.

For S finite, it turns out that there are still sufficiently many periodic orbits to have the growth rate result (10), but the asymptotic (8) is modified in much the same way as Waddington observed for quasi-hyperbolic toral automorphisms:

$$\liminf_{X \rightarrow \infty} \frac{X \pi_T(X)}{e^{(X+1)h_{\text{top}}(T)}} > 0 \quad (9)$$

and there is an associated pair (X^*, a_T) , where X^* is a compact group and $a_T \in X^*$, with the property that if $a_T^{N_j}$ converges in X^* as $j \rightarrow \infty$, then there is convergence in (9).

A simple special case will illustrate this. Taking $\mathbb{K} = \mathbb{Q}$, $\xi = 2$, and $S = \{3\}$ gives a compact group endomorphism T with

$$F_n(T) = (2^n - 1) |2^n - 1|_3.$$

For this example, the results of Chothi et al. (1997) are sharper: The expression in (9) converges along (X_j) if and only if 2^{X_j} converges in the ring of 3-adic integers $\mathbb{Z}_3$, the expression has uncountably many limit points, and the upper and lower limits are transcendental.

Similarly, the dynamical analogue of Mertens' theorem found by Sharp may be found for S -integer systems with S finite. Writing

$$\mathcal{M}_T(N) = \sum_{|\tau| \leq N} \frac{1}{e^{h(T)|\tau|}},$$

it is shown in Chothi et al. (1997) that for an ergodic S -integer map T with $\mathbb{K} = \mathbb{Q}$ and S finite, there are constants $k_T \in \mathbb{Q}$ and C_T such that

$$\mathcal{M}_T(N) = k_T \log N + C_T + O(1/N).$$

Without the restriction that $\mathbb{K} = \mathbb{Q}$, it is shown that there are constants $k_T \in \mathbb{Q}$, C_T , and $\delta > 0$ with

$$\mathcal{M}_T(N) = k_T \log N + C_T + O(N^{-\delta}).$$

The case of S co-finite (that is, containing all but finitely many primes) gives rise to a large class of group automorphisms with a polynomial bound on the growth of the number of closed periodic orbits. The natural analytical tool here is the *orbit Dirichlet series*

$$d_T(z) = \sum_{n \geq 1} \frac{O_T(n)}{n^z},$$

and a suitable form of 'Dirichlet rationality' and relations between its analytical behaviour near its abscissa of convergence and orbit growth properties are shown in work of Everest et al. (2010).

Exotic Orbit Growth

The results above raise the question of what the possible orbit growth phenomena are for compact group automorphisms. A consequence of Linnik's bounds for first appearance of primes in arithmetic progressions (Linnik 1944) can be used (Ward 2005) to prove that for any $C \in [0, \infty]$ there is a compact group automorphism T with the property that

$$\frac{1}{n} \log F_n(T) \rightarrow C$$

as $n \rightarrow \infty$. This construction was further refined by Haynes and White (2014) to exhibit other possible growth rates. The number-theoretic questions thrown up here are subtle ones, and the constructions are unsatisfying in that the underlying compact group is totally disconnected and the group automorphism is not ergodic with respect to

Haar measure. On connected groups, the S -integer construction has been used by Baier et al. (2013) to show that the following “exotic” orbit growth properties occur.

- For any $\kappa \in (0, 1)$, there is an ergodic compact connected group automorphism $T: X \rightarrow X$ with $\mathcal{M}_T(N) \sim \kappa \log N$.
- For any $r \in \mathbb{N}$ and $\kappa > 0$, there is an ergodic compact connected group automorphism $T: X \rightarrow X$ with $\mathcal{M}_T(N) \sim \kappa(\log \log N)^r$.
- For any $\delta \in (0, 1)$ and $k > 0$, there is an ergodic compact connected group automorphism $T: X \rightarrow X$ with $\mathcal{M}_T(N) \sim k(\log N)^\delta$.

We refer to the survey of Miles et al. (2015) for an overview.

Pólya–Carlson Dichotomy

Everest et al. (2005) noticed that the very simplest example of a nontrivial S -integer system, namely, the automorphism dual to $x \mapsto 2x$ on $\mathbb{Z}[\frac{1}{6}]$, had the property that its dynamical zeta function admitted a natural boundary at its circle of convergence. This phenomenon later emerged as a pervasive feature of group automorphisms, and Bell et al. (2014) formulated the conjecture that compact group automorphisms exhibit a Pólya–Carlson dichotomy: their dynamical zeta function is either rational or admits a natural boundary. This has been proved in many cases with S or its complement finite, but in general remains open. This property certainly does not hold for maps in general, as it is easy to find examples of maps whose dynamical zeta function is a nonrational function satisfying an algebraic identity.

Similar questions arise for endomorphisms of abelian varieties, and in positive characteristic, the relationship between rationality, algebraicity, and the existence of natural boundaries for the dynamical zeta function has been studied by Byszewski and Cornelissen (2018), where once again a dichotomy of Pólya–Carlson type is found.

Diophantine Analysis as a Toolbox

Many problems in ergodic theory and dynamical system exploit ideas and results from number

theory in a direct way; we illustrate this by describing a selection of dynamical problems that call on particular parts of number theory in an essential way. The example of mixing in section “Mixing and Additive Relations in Fields” is particularly striking for two reasons: the results needed from number theory are relatively recent, and the ergodic application directly motivated a further development in number theory.

Orbit Growth and Convergence

The analysis of periodic orbits – how their number grows as the length grows and how they spread out through space – is of central importance in dynamics (see Katok (1980) for example). An instance of this is that for many simple kinds of dynamical systems $T: X \rightarrow X$ (where T is a continuous map of a compact metric space (X, d)), the logarithmic growth rate of the number of periodic points exists and coincides with the topological entropy $h(T)$ (an invariant giving a quantitative measure of the average rate of growth in orbit complexity under T). That is,

$$\frac{1}{n} \log F_n(T) \rightarrow h_{\text{top}}(T) \quad (10)$$

for many of the simplest dynamical systems. For example, if $X = \mathbb{T}^r$ is the r -torus and $T = T_A$ is the automorphism of the torus corresponding to a matrix A in $GL_r(\mathbb{Z})$, then T_A is ergodic with respect to Lebesgue measure if and only if no eigenvalue of A is a root of unity. Under this assumption, we have

$$F_n(T_A) = \prod_{i=1}^r |\lambda_i^n - 1|$$

and

$$h_{\text{top}}(T_A) = \sum_{i=1}^r \log \max \{1, |\lambda_i|\} \quad (11)$$

where $\lambda_1, \dots, \lambda_r$ are the eigenvalues of A . It follows that the convergence in (10) is clear under the assumption that T_A is *hyperbolic* (that is, no eigenvalue has modulus one). Without this

assumption, the convergence is less clear: for $r \geq 4$ the automorphism T_A may be ergodic without being hyperbolic. That is, while no eigenvalues are unit roots some may have unit modulus. As pointed out by Lind (1982) in his study of these *quasihyperbolic* automorphisms, the convergence (10) does still hold for these systems, but this requires a significant Diophantine result (the theorem of Gel'fond (1960) suffices; one may also use Baker's theorem (Baker 1990)). Even further from hyperbolicity lie the family of S -integer systems (Chothi et al. 1997; Ward 1998); their orbit-growth properties are intimately tied up with Artin's conjecture on primitive roots and prime divisors of linear recurrence sequences.

Mixing and Additive Relations in Fields

The problem of higher-order mixing for commuting group automorphisms provides a striking example of the dialogue between ergodic theory and number theory, in which deep results from number theory have been used to solve problems in ergodic theory, and questions arising in ergodic theory have motivated further developments in number theory.

An action T of a countable group Γ on a probability space $(X, \mathcal{B}, \mu)$ is called *k-fold mixing* or *mixing on $(k+1)$ sets* if

$$\mu(A_0 \cap T_{-g_1}A_1 \cap \dots \cap T_{-g_k}A_k) \rightarrow \mu(A_0) \cdots \mu(A_k) \quad (12)$$

as

$$g_i g_j^{-1} \rightarrow \infty \text{ for } i \neq j$$

with the convention that $g_0 = 1_\Gamma$, for any sets $A_0, \dots, A_k \in \mathcal{B}$; $g_n \rightarrow \infty$ in Γ means that for any finite set $F \subset \Gamma$ there is an N with $n > N \Rightarrow g_n \notin F$. For $k = 1$ the property is called simply *mixing*. This notion for single transformations goes back to the foundational work of Rohlin (1949), where he showed that ergodic group endomorphisms are mixing of all orders (and so the notion is not useful for distinguishing between group endomorphisms as measurable dynamical systems). He

raised the (still open) question of whether any measure-preserving transformation can be mixing without being mixing of all orders.

A class of group actions that are particularly easy to understand are the *algebraic dynamical systems* studied systematically by Schmidt (1995): here X is a compact abelian group, each T_g is a continuous automorphism of X , and μ is the Haar measure on X . Schmidt (1989) related mixing properties of algebraic dynamical systems with $\Gamma = \mathbb{Z}^d$ to statements in arithmetic and showed that a mixing action on a connected group could only fail to mix in a certain way. Later Schmidt and the author (Schmidt and Ward 1993) showed that for X connected, mixing implies mixing of all orders. The proof proceeds by showing that the result is exactly equivalent to the following statement: if $\mathbb{K}$ is a field of characteristic zero, and G is a finitely generated subgroup of the multiplicative group $\mathbb{K}^\times$, then the equation

$$a_1 x_1 + \dots + a_n x_n = 1 \quad (13)$$

for fixed $a_1, \dots, a_n \in \mathbb{K}^\times$ has a finite number of solutions $x_1, \dots, x_n \in G$ for which no subsum $\sum_{i \in I} a_i x_i$ with $I \subsetneq \{1, \dots, n\}$ vanishes. The bound on the number of solutions to (13) follows from the profound extensions to W. Schmidt's subspace theorem in Diophantine geometry (Schmidt 1972) by Evertse and Schlickewei (see Evertse and Schlickewei (1999), van der Poorten and Schlickewei (1991), and Schlickewei (1990)) for the details).

The argument in Schmidt and Ward (1993) may be cast as follows: failure of k -fold mixing in a connected algebraic dynamical system implies (via duality) an infinite set of solutions to an equation of the shape (13) in some field of characteristic zero. The S -unit theorem means that this can only happen if there is some proper subsum that vanishes infinitely often. This infinite family of solutions to a homogeneous form of (13) with fewer terms can then be translated back via duality to show that the system fails to mix for some strictly lower order, proving that mixing implies mixing of all orders by induction.

Stronger Diophantine results allow these arguments to be extended in various directions. Miles

and Ward (2006) used explicit bounds on the number of solutions of S -integer systems due to Evertse et al. (2002) to show that mixing implies mixing of all orders for actions of $\mathbb{Q}^d$ by automorphisms of a compact connected abelian group.

Mixing properties for algebraic dynamical systems without the assumption of connectedness are quite different, and in particular, it is possible to have mixing actions that are not mixing of all orders. This is a simple consequence of the fact that the constituents of a disconnected algebraic dynamical system are associated with fields of positive characteristic, where the presence of the Frobenius automorphism can prevent higher-order mixing. Ledrappier (1978) pointed this out via examples of the following shape. Let

$$X = \left\{ x \in \mathbb{F}_2^{\mathbb{Z}^2} : x_{(a+1,b)} + x_{(a,b)} + x_{(a,b+1)} = 0 \pmod{2} \right\}$$

and define the $\mathbb{Z}^2$ -action T to be the natural shift action,

$$(T_{(n,m)}x)_{(a,b)} = x_{(a+n,b+m)}.$$

It is readily seen that this action is mixing with respect to the Haar measure. The condition $x_{(a+1,b)} + x_{(a,b)} + x_{(a,b+1)} = 0 \pmod{2}$ implies that, for any $k \geq 1$,

$$\begin{aligned} x_{(0,2^k)} &= \sum_{j=0}^{2^k} \binom{2^k}{j} x_{(j,0)} \\ &= x_{(0,0)} + x_{(2^k,0)} \pmod{2} \end{aligned} \quad (14)$$

since every entry in the 2^k th row of Pascal's triangle is even apart from the first and the last. Now let $A = \{x \in X : x_{(0,0)} = 0\}$ and let $x_* \in X$ be any element with $x_{(0,0)} = 1$. Then X is the disjoint union of A and $A + x_*$, so

$$\mu(A) = \mu(A + x_*) = \frac{1}{2}.$$

However, (14) shows that

$$x \in A \cap T_{-(2^k,0)}A \Rightarrow x \in T_{-(0,2^k)}A,$$

so

$$A \cap T_{-(2^k,0)}A \cap T_{-(0,2^k)}(A + x_*) = \emptyset$$

for all $k \geq 1$, which shows that T cannot be mixing on three sets. Arenas-Carmona et al. (2008) showed that if the dilates of the specific shape $\{(0, 0), (1, 0), (0, 1)\}$ are avoided, then Ledrappier's system exhibits mixing of all orders.

The full picture of higher-order mixing properties on disconnected groups is rather involved; see Schmidt's monograph (Schmidt 1995). A simple illustration is the construction by Einsiedler and the author (Einsiedler and Ward 2003) of systems with any prescribed order of mixing. When such systems fail to be mixing of all orders, they fail in a very specific way – along dilates of a specific *shape* (a finite subset of $\mathbb{Z}^d$). One insight into the range of possible behaviors for algebraic dynamical systems of this is seen in the following result. Call a shape “admissible” if it does not lie on a line in $\mathbb{Z}^d$, does contain 0, and has the property that for any $k > 1$ the set $\frac{1}{k}S$ contains nonintegral points. Then it is shown in Ward (1997) that if S and T are admissible shapes, then there is an algebraic $\mathbb{Z}^d$ -action that is mixing on S and not mixing on T unless a translate of T is a subset of S . In Ledrappier's example above, the shape that fails to mix is $\{(0, 0), (1, 0), (0, 1)\}$. This gives an order of mixing as detected by shapes; computing this is in principle an *algebraic* problem. On the other hand, there is a more natural definition of the order of mixing, namely, the largest k for which (12) holds; computing this is in principle a *Diophantine* problem. A conjecture emerged (formulated explicitly by Schmidt (2001)) that for any algebraic dynamical system, if every set of cardinality $r \geq 2$ is a mixing shape, then the system is mixing on r sets.

This question motivated Masser (2004) to prove an appropriate analogue of the S -unit theorem on the number of solutions to (13) in positive characteristic as follows. Let H be a multiplicative group and fix $n \in \mathbb{N}$. An infinite subset $A \subset H^n$ is called *broad* if it has both of the following properties:

- if $h \in H$ and $1 \leq j \leq n$, then there are at most finitely many $(a_1, \dots, a_n)$ in A with $a_j = h$;
- if $n \geq 2$, $h \in H$ and $1 \leq i < j \leq n$ then there are at most finitely many $(a_1, \dots, a_n) \in H$ with $a_i a_j^{-1} = h$.

Then Masser's theorem says the following. Let $\mathbb{K}$ be a field of characteristic $p > 0$, let G be a finitely generated subgroup of $\mathbb{K}^\times$ and suppose that the equation

$$a_1 x_1 + \dots + a_n x_n = 1$$

has a broad set of solutions $(x_1, \dots, x_n) \in G^n$ for some constants $a_1, \dots, a_n \in \mathbb{K}^\times$. Then there is an $m \leq n$, constants $b_1, \dots, b_m \in \mathbb{K}^\times$ and some $(g_1, \dots, g_m) \in G^m$ with the following properties:

- $g_j \neq 1$ for $1 \leq j \leq m$;
- $g_i g_j^{-1} \neq 1$ for $1 \leq i < j \leq m$;
- there are infinitely many k for which

$$b_1 g_1^k + b_2 g_2^k + \dots + b_m g_m^k = 1.$$

The proof that shapes detect the order of mixing in algebraic dynamics then proceeds much as in the connected case.

The extent to which all this behavior is a special feature of *algebraic* dynamical systems is illustrated by a construction in Ward (1997) of a measure-preserving $\mathbb{Z}^d$ -action T on a probability space $(X, \mathcal{B}, \mu)$ for any $d \geq 2$ with the following two properties:

- T is *rigid*, meaning that there is a sequence $(\mathbf{n}_j)$ going to infinity in $\mathbb{Z}^d$ as $j \rightarrow \infty$ with the property that $\mu(A \cap T_{\mathbf{n}_j} A) \rightarrow \mu(A)$ as $j \rightarrow \infty$ for all $A \in \mathcal{B}$ (an extreme form of non-mixing);
- Every shape $S \subset \mathbb{Z}^d$ is mixing for T .

The construction uses ideas of Ferenczi and Kamiński (1995) and is built as a Gaussian dynamical system.

A natural question for mixing algebraic dynamical systems that are mixing concerns the *rate* at which the mixing occurs. That is, for a suitable class $\mathcal{S}$ of functions $X \rightarrow \mathbb{C}$ with a natural norm $\|\cdot\|_{\mathcal{S}}$ on a compact group X carrying a $\mathbb{Z}^d$ -action T that is known to be mixing of all orders, what can be said for each $k \in \mathbb{N}$ about functions Φ_k satisfying

$$\left| \int_X f_1(T_{\mathbf{n}_1} x) \cdots f_k(T_{\mathbf{n}_k} x) d\mu(x) - \int_X f_1 d\mu \cdots \int_X f_k d\mu \right| \leq \Phi_k \left(\min_{1 \leq i < j \leq k} \{\|\mathbf{n}_i - \mathbf{n}_j\|\} \right) \prod_{i=1}^k \|f_i\|_{\mathcal{S}}$$

for all $f_1, \dots, f_k \in \mathcal{S}$? Mixing of all orders means there is such a function satisfying $\Phi_k(n) \rightarrow 0$ as $n \rightarrow \infty$ for $\mathcal{S} = L^\infty(X)$, and a “rate of mixing” result is any improvement on this, typically for a much more restricted class of functions. Miles and the author (Waldschmidt 1993) used Baker's theorem and its p -adic variants to establish a very weak general result of this shape under the assumption that X has finite topological dimension, but in a form that makes it extremely difficult to be explicit about Φ_k even in concrete examples. Using more delicate Diophantine estimates for linear forms in logarithms of algebraic numbers due to Waldschmidt (1993), Gorodnik and Spatzier (2015) found explicit rates of mixing for $k = 3$ in the setting of nilmanifolds, a new result even in the toral case. To state their main result, let T be a $\mathbb{Z}^d$ -action by automorphisms of a compact nilmanifold $X = G/\Gamma$ with the property that each transformation $T_{\mathbf{n}}$ for $\mathbf{n} \neq 0$ is ergodic with respect to the measure μ induced by Haar measure on G . Then they construct a function $\theta \mapsto \eta(\theta) > 0$ such that if $C^\theta(X)$ denotes the class of Hölder functions on X with exponent θ and Hölder norm $\|\cdot\|_\theta$, then there is a constant $C > 0$ such that

$$\left| \int f(T_{\mathbf{l}} x) g(T_{\mathbf{m}} x) h(T_{\mathbf{n}} x) d\mu(x) - \int f d\mu \int g d\mu \int h d\mu \right| \leq C (\exp(\min\{\|\mathbf{l} - \mathbf{m}\|, \|\mathbf{l} - \mathbf{n}\|, \|\mathbf{n} - \mathbf{m}\|\}))^{-\eta(\theta)} \|f\|_\theta \|g\|_\theta \|h\|_\theta$$

for all $f, g, h \in C^\theta(X)$.

Future Directions

The interaction between ergodic theory, number theory, and combinatorics continues to expand

rapidly, and many future directions of research are discussed in the entries ► [“Ergodic Theory on Homogeneous Spaces and Metric Number Theory”](#) by Kleinbock, ► [“Ergodic Theory: Rigidity”](#) by Nițică, and ► [“Ergodic Theory: Recurrence”](#) by Frantzikinakis and McCutcheon. Some of the directions most relevant to the examples discussed in this entry include the following.

The recent developments mentioned in section [“Sets of Primes”](#) clearly open many exciting prospects involving finding new structures in arithmetically significant sets (like the primes). The original conjecture of Erdős and Turán (1936) asked if $\sum_{a \in A \subset \mathbb{N}} \frac{1}{a} = \infty$ is sufficient to force the set A to contain arbitrary long arithmetic progressions and remains open. This would of course imply both Szemerédi’s theorem (Szemerédi 1975) and the result of Green and Tao (2008) on arithmetic progressions in the primes. More generally, it is clear that there is still much to come from the dialogue subsuming the four parallel proofs of Szemerédi’s: one by purely combinatorial methods, one by ergodic theory, one by hypergraph theory, and one by Fourier analysis and additive combinatorics. For an overview, see the survey papers of Tao (2006, 2007a, b).

In the context of the orbit-counting results in section [“Orbit-Counting as an Analogous Development,”](#) a natural problem is to, on the one hand, obtain finer asymptotics with better control of the error terms and, on the other hand, to extend the situations that can be handled. In particular, relaxing the hypotheses related to hyperbolicity (or negative curvature) is a constant challenge.

The rate of mixing result at the end of section [“Mixing and Additive Relations in Fields”](#) points towards a large array of natural conjectures concerned with rate of mixing, central limit theorem phenomena, and rigidity of cocycles for $\mathbb{Z}^d$ -actions.

Bibliography

Primary Literature

Adler RL, Weiss B (1970) Similarity of automorphisms of the torus. *Memoirs of the American Mathematical Society*, No. 98. American Mathematical Society, Providence

- Anosov DV (1967) Geodesic flows on closed Riemannian manifolds of negative curvature. *Trudy Mat Inst Steklov* 90:209
- Arenas-Carmona L, Berend D, Bergelson V (2008) Ledrappier’s system is almost mixing of all orders. *Ergodic Theory Dynam Systems* 28(2):339–365
- Auslander L, Green L, Hahn F (1963) Flows on homogeneous spaces. With the assistance of L. Markus and W. Massey, and an appendix by L. Greenberg. *Annals of Mathematics Studies*, No. 53. Princeton University Press, Princeton
- Baier S, Jaidee S, Stevens S, Ward T (2013) Automorphisms with exotic orbit growth. *Acta Arith* 158(2): 173–197
- Baker A (1990) *Transcendental number theory*, 2nd edn. Cambridge mathematical library. Cambridge University Press, Cambridge
- Bell J, Miles R, Ward T (2014) Towards a Pólya–Carlson dichotomy for algebraic dynamics. *Indag Math (NS)* 25(4):652–668
- Benford F (1938) The law of anomalous numbers. *Proc Am Philos Soc* 78:551–572
- Bergelson V (2003) Minimal idempotents and ergodic Ramsey theory. In: *Topics in dynamics and ergodic theory*, volume 310 of *London Math. Soc. Lecture Note Ser.* Cambridge University Press, Cambridge, pp 8–39
- Bergelson V, Leibman A (1996) Polynomial extensions of van der Waerden’s and Szemerédi’s theorems. *J Am Math Soc* 9(3):725–753
- Bergelson V, McCutcheon R (2000) An ergodic IP polynomial Szemerédi theorem. *Mem Amer Math Soc* 146(695):viii+106
- Birkhoff GD (1931) Proof of the ergodic theorem. *Proc Natl Acad Sci U S A* 17:656–660
- Bohl P (1909) Über ein in der Theorie der säkularen Störungen vorkommendes Problem. *J Undergrad Math* 135:189–283
- Borel E (1909) Les probabilités denombrables et leurs applications arithmétiques. *Rend Circ Math Palermo* 27:247–271
- Bourgain J (1988a) An approach to pointwise ergodic theorems. In: *Geometric aspects of functional analysis* (1986/87), volume 1317 of *Lecture Notes in Math.* Springer, Berlin, pp 204–223
- Bourgain J (1988b) On the maximal ergodic theorem for certain subsets of the integers. *Israel J Math* 61(1): 39–72
- Bowen R (1970) Markov partitions for axiom A diffeomorphisms. *Am J Math* 92:725–747
- Bowen R (1972) The equidistribution of closed geodesics. *Am J Math* 94:413–423
- Bowen R (1973) Symbolic dynamics for hyperbolic flows. *Am J Math* 95:429–460
- Byszewski J, Cornelissen G (2018) Dynamics on abelian varieties in positive characteristic. *Algebra Number Theory* 12(9):2185–2235
- Chothi V, Everest G, Ward T (1997) S-integer dynamical systems: periodic points. *J Reine Angew Math* (489): 99–132

- Dolgopyat D (1998) On decay of correlations in Anosov flows. *Ann Math* (2) 147(2):357–390
- Einsiedler M, Ward T (2003) Asymptotic geometry of non-mixing sequences. *Ergod Theor Dyn Syst* 23(1):75–85
- Erdős P (1949) On a new method in elementary number theory which leads to an elementary proof of the prime number theorem. *Proc Nat Acad Sci U S A* 35:374–384
- Erdős P, Turán P (1936) On some sequences of integers. *J Lond Math Soc* 11:261–264
- Everest G, Stangoe V, Ward T (2005) Orbit counting with an isometric direction. In: *Algebraic and topological dynamics*, volume 385 of *Contemp. Math.* American Mathematical Society, Providence, pp 293–302
- Everest G, Miles R, Stevens S, Ward T (2007) Orbit-counting in non-hyperbolic dynamical systems. *J Reine Angew Math* 608:155–182
- Everest G, Miles R, Stevens S, Ward T (2010) Dirichlet series for finite combinatorial rank dynamics. *Trans Am Math Soc* 362(1):199–227
- Evertse JH, Schlickewei HP (1999) The absolute subspace theorem and linear equations with unknowns from a multiplicative group. In: *Number theory in progress*, Vol. 1 (Zakopane-Kościełisko, 1997). de Gruyter, Berlin, pp 121–142
- Evertse J-H, Schlickewei HP, Schmidt WM (2002) Linear equations in variables which lie in a multiplicative group. *Ann Math* (2) 155(3):807–836
- Ferenczi S, Kamiński B (1995) Zero entropy and directional Bernoullicity of a Gaussian $\mathbb{Z}^2$ -action. *Proc Am Math Soc* 123(10):3079–3083
- Furstenberg H (1961) Strict ergodicity and transformation of the torus. *Am J Math* 83:573–601
- Furstenberg H (1977) Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J Analyse Math* 31:204–256
- Furstenberg H, Katznelson Y (1978) An ergodic Szemerédi theorem for commuting transformations. *J Analyse Math* 34:275–291. (1979)
- Furstenberg H, Katznelson Y (1985) An ergodic Szemerédi theorem for IP-systems and combinatorial theory. *J Analyse Math* 45:117–168
- Furstenberg H, Weiss B (1978) Topological dynamics and combinatorial number theory. *J Analyse Math* 34:61–85. (1979)
- Gel'fond AO (1960) *Transcendental and algebraic numbers*. Translated from the first Russian edition by Leo F. Boron. Dover, New York
- Goldfeld D (2004) The elementary proof of the prime number theorem: an historical perspective. In: *Number theory* (New York, 2003). Springer, New York, pp 179–192
- Goldston DA, Pintz J, Yıldırım CY (2009) Primes in tuples. I. *Ann Math* (2) 170(2):819–862
- Gorodnik A, Spatzier R (2015) Mixing properties of commuting nilmanifold automorphisms. *Acta Math* 215(1):127–159
- Gowers WT (2007) Hypergraph regularity and the multi-dimensional Szemerédi theorem. *Ann Math* (2) 166(3):897–946
- Green B, Tao T (2008) The primes contain arbitrarily long arithmetic progressions. *Ann Math* (2) 167(2):481–547
- Green B, Tao T (2010) Linear equations in primes. *Ann Math* 171(3):1753–1850
- Haynes A, White CJ (2014) Group automorphisms with prescribed growth of periodic points, and small primes in arithmetic progressions in intervals. *Adv Math* 252:572–585
- Hill TP (1995) Base-invariance implies Benford's law. *Proc Am Math Soc* 123(3):887–895
- Hindman N (1974) Finite sums from sequences within cells of a partition of $\mathbb{N}$. *J Combin Theory Ser A* 17:1–11
- Hlawka E (1964) Discrepancy and uniform distribution of sequences. *Compos Math* 16(83–91):1964
- Host B, Kra B (2005a) Convergence of polynomial ergodic averages. *Israel J Math* 149:1–19. *Probability in mathematics*
- Host B, Kra B (2005b) Nonconventional ergodic averages and nilmanifolds. *Ann Math* (2) 161(1):397–488
- Huber H (1959) *Zur analytischen Theorie hyperbolischen Raumformen und Bewegungsgruppen*. *Math Ann* 138:1–26
- Katok A (1980) Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst Hautes Études Sci Publ Math* 51:137–173
- Katsuda A, Sunada T (1990) Closed orbits in homology classes. *Inst Hautes Études Sci Publ Math* 71:5–32
- Khanin K, Lopes Dias J, Marklof J (2007) Multi-dimensional continued fractions, dynamical renormalization and KAM theory. *Commun Math Phys* 270(1):197–231
- Khinchin AI (1964) *Continued fractions*. The University of Chicago Press, Chicago/London
- Knieper G (1997) On the asymptotic geometry of non-positively curved manifolds. *Geom Funct Anal* 7(4):755–782
- Koopman B (1931) Hamiltonian systems and transformations in Hilbert spaces. *Proc Natl Acad Sci U S A* 17:315–318
- Kuz'min RO (1928) A problem of Gauss. *Dokl Akad Nauk* 1928:375–380
- Lagarias JC (1994) Geodesic multidimensional continued fractions. *Proc Lond Math Soc* (3) 69(3):464–488
- Lalley SP (1987) Distribution of periodic orbits of symbolic and axiom a flows. *Adv Appl Math* 8(2):154–193
- Lalley SP (1988) The “prime number theorem” for the periodic orbits of a Bernoulli flow. *Am Math Mon* 95(5):385–398
- Ledrappier F (1978) Un champ markovien peut être d'entropie nulle et mélangeant. *C R Acad Sci Paris Ser A-B* 287(7):A561–A563
- Leibman A (2005) Convergence of multiple ergodic averages along polynomials of several variables. *Israel J Math* 146:303–315
- Levy P (1929) Sur les lois de probabilité dont dépendent les quotients complets et incomplets d'une fraction continue. *Bull Soc Math France* 57:178–194
- Lévy P (1936) Sur quelques points de la théorie des probabilités dénombrables. *Ann Inst H Poincaré* 6(2):153–184
- Lind DA (1982) Dynamical properties of quasihyperbolic toral automorphisms. *Ergod Theor Dyn Syst* 2(1):49–68

- Linnik UV (1944) On the least prime in an arithmetic progression. I. The basic theorem. *Rec Math [Mat Sbornik]* NS 15(57):139–178
- Margulis GA (1969) Certain applications of ergodic theory to the investigation of manifolds of negative curvature. *Funkcional Anal i Priložen* 3(4):89–90
- Margulis GA (2004) On some aspects of the theory of Anosov systems. *Springer Monographs in Mathematics*. Springer, Berlin. With a survey by Richard Sharp: Periodic orbits of hyperbolic flows, Translated from the Russian by Valentina Vladimirovna Szulikowska
- Masser DW (2004) Mixing and linear equations over groups in positive characteristic. *Israel J Math* 142: 189–204
- Mertens F (1874) Ein Beitrag zur analytischen Zahlentheorie. *J Reine Angew Math* 78:46–62
- Miles R, Ward T (2006) Mixing actions of the rationals. *Ergod Theor Dyn Syst* 26(6):1905–1911
- Miles R, Ward T (2011) A directional uniformity of periodic point distribution and mixing. *Discrete Contin Dyn Syst* 30(4):1181–1189
- Nagle B, Rödl V, Schacht M (2006) The counting lemma for regular k -uniform hypergraphs. *Random Struct Algorith* 28(2):113–179
- Newcomb S (1881) Note on the frequency of the use of digits in natural numbers. *Am J Math* 4(1):39–40
- Noorani MSM (1999) Mertens' theorem and closed orbits of ergodic toral automorphisms. *Bull Malaysian Math Soc* (2) 22(2):127–133
- Oxtoby JC (1952) Ergodic sets. *Bull Am Math Soc* 58: 116–136
- Parry W (1969) Ergodic properties of affine transformations and flows on nilmanifolds. *Am J Math* 91: 757–771
- Parry W (1983) An analogue of the prime number theorem for closed orbits of shifts of finite type and their suspensions. *Israel J Math* 45(1):41–52
- Parry W (1984) Bowen's equidistribution theory and the Dirichlet density theorem. *Ergod Theor Dyn Syst* 4(1): 117–134
- Parry W, Pollicott M (1983) An analogue of the prime number theorem for closed orbits of axiom A flows. *Ann Math* 118(3):573–591
- Poincaré H (1890) Sur le problème des trois corps et les équations de la Dynamique. *Acta Math* 13:1–270
- Pollicott M, Sharp R (1998) Exponential error terms for growth functions on negatively curved surfaces. *Am J Math* 120(5):1019–1042
- Rado R (1933) Studien zur Kombinatorik. *Math Z* 36(1): 424–470
- Ratner M (1973) Markov partitions for Anosov flows on n -dimensional manifolds. *Israel J Math* 15:92–114
- Rohlin VA (1949) On endomorphisms of compact commutative groups. *Izvestiya Akad Nauk SSSR Ser Mat* 13:329–340
- Roth K (1952) Sur quelques ensembles d'entiers. *C R Acad Sci Paris* 234:388–390
- Sárközy A (1978) On difference sets of sequences of integers. III. *Acta Math Acad Sci Hungar* 31(3–4): 355–386
- Schlickewei HP (1990) S-unit equations over number fields. *Invent Math* 102(1):95–107
- Schmidt WM (1972) Norm form equations. *Ann Math* 96:526–551
- Schmidt K (1989) Mixing automorphisms of compact groups and a theorem by Kurt Mahler. *Pac J Math* 137(2):371–385
- Schmidt K, Ward T (1993) Mixing automorphisms of compact groups and a theorem of Schlickewei. *Invent Math* 111(1):69–76
- Selberg A (1949) An elementary proof of the prime-number theorem. *Ann Math* 50(2):305–313
- Selberg A (1956) Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. *J Indian Math Soc (NS)* 20:47–87
- Sharp R (1991) An analogue of Mertens' theorem for closed orbits of axiom A flows. *Bol Soc Brasil Mat (NS)* 21(2):205–229
- Sharp R (1993) Closed orbits in homology classes for Anosov flows. *Ergod Theor Dyn Syst* 13(2):387–408
- Sierpiński W (1910) Sur la valeur asymptotique d'une certaine somme. *Bull Int Acad Pol Sci Lett (Cracovie)* 1910:9–11
- Sinaĭ JG (1966) Asymptotic behavior of closed geodesics on compact manifolds with negative curvature. *Izv Akad Nauk SSSR Ser Mat* 30:1275–1296
- Sinaĭ JG (1968) Construction of Markov partitions. *Funkcional Anal i Priložen* 2(3):70–80
- Smale S (1967) Differentiable dynamical systems. *Bull Am Math Soc* 73:747–817
- Szemerédi E (1969) On sets of integers containing no four elements in arithmetic progression. *Acta Math Acad Sci Hungar* 20:89–104
- Szemerédi E (1975) On sets of integers containing no k elements in arithmetic progression. *Acta Arith* 27:199–245
- Tao T, Ziegler T (2008) The primes contain arbitrarily long polynomial progressions. *Acta Math* 201(2):213–305
- van der Poorten AJ, Schlickewei HP (1991) Additive relations in fields. *J Austral Math Soc Ser A* 51(1):154–170
- van der Waerden BL (1927) Beweis einer Baudet'schen Vermutung. *Nieuw Arch Wisk* 15:212–216
- von Neumann J (1932) Proof of the quasi-ergodic hypothesis. *Proc Natl Acad Sci U S A* 18:70–82
- Waddington S (1991) The prime orbit theorem for quasi-hyperbolic toral automorphisms. *Monatsh Math* 112(3):235–248
- Waldschmidt M (1993) Minorations de combinaisons linéaires de logarithmes de nombres algébriques. *Can J Math* 45(1):176–224
- Ward T (1997/98) Three results on mixing shapes. *N Y J Math* 3A (Proceedings of the New York Journal of Mathematics Conference, June 9–13, 1997):1–10
- Ward T (1998) Almost all S-integer dynamical systems have many periodic points. *Ergod Theor Dyn Syst* 18(2):471–486
- Ward T (2005) Group automorphisms with few and with many periodic points. *Proc Am Math Soc* 133(1): 91–96

- Weyl H (1910) Über die *Gibbssche* Erscheinung und verwandte Konvergenzphänomene. *Rendiconti del Circolo Matematico di Palermo* 30:377–407
- Weyl H (1916) Über die Gleichverteilung von Zahlen mod Eins. *Math Ann* 77:313–352
- Wiener N (1932) Tauberian theorems. *Ann Math* 33(1): 1–100

Books and Reviews

- Arnol'd VI, Avez A (1968) Ergodic problems of classical mechanics. Translated from the French by A. Avez. W. A. Benjamin, New York/Amsterdam
- Bergelson V (1996) Ergodic Ramsey theory – an update. In: *Ergodic theory of $\mathbb{Z}^d$ actions* (Warwick, 1993–1994), volume 228 of *London Math. Soc. Lecture Note Ser.* Cambridge University Press, Cambridge, pp 1–61
- Bergelson V (2000) Ergodic theory and Diophantine problems. In: *Topics in symbolic dynamics and applications* (Temuco, 1997), volume 279 of *London Math. Soc. Lecture Note Ser.* Cambridge University Press, Cambridge, pp 167–205
- Bergelson V (2006) Combinatorial and Diophantine applications of ergodic theory. In: *Handbook of dynamical systems*, vol 1B. Elsevier B. V., Amsterdam, pp 745–869. Appendix A by A. Leibman and Appendix B by Anthony Quas and Máté Wierdl
- Cornfeld IP, Fomin SV, Sinai YG (1982) *Ergodic theory*. Springer, New York
- Dajani K, Kraaikamp C (2002) *Ergodic theory of numbers*, volume 29 of *Carus mathematical monographs*. Mathematical Association of America, Washington, DC
- del Junco A (2008) Ergodic theorems. In: *Encyclopedia of complexity and systems science*. Springer
- Denker M, Grillenberger C, Sigmund K (1976) *Ergodic theory on compact spaces*. *Lecture notes in mathematics*, vol 527. Springer, Berlin
- Einsiedler M, Ward T (2011) *Ergodic theory with a view towards number theory*, volume 259 of *Graduate texts in mathematics*. Springer, London
- Ellis R (1969) *Lectures on topological dynamics*. W. A. Benjamin, New York
- Furstenberg H (1981) *Recurrence in ergodic theory and combinatorial number theory*. Princeton University Press, Princeton. M. B. Porter Lectures
- Furstenberg H, Katznelson Y, Ornstein D (1982) The ergodic theoretical proof of Szemerédi's theorem. *Bull Am Math Soc (NS)* 7(3):527–552
- Glasner E (2003) *Ergodic theory via joinings*, volume 101 of *Mathematical surveys and monographs*. American Mathematical Society, Providence
- Hejhal DA (1976) The Selberg trace formula and the Riemann zeta function. *Duke Math J* 43(3):441–482
- Iosifescu M, Kraaikamp C (2002) *Metrical theory of continued fractions*, volume 547 of *Mathematics and its applications*. Kluwer Academic, Dordrecht
- Katok A, Hasselblatt B (1995) *Introduction to the modern theory of dynamical systems*, volume 54 of *Encyclopedia of mathematics and its applications*. Cambridge University Press, Cambridge. With a supplementary chapter by Anatole Katok and Leonardo Mendoza
- Kra B (2006) The Green-Tao theorem on arithmetic progressions in the primes: an ergodic point of view. *Bull Am Math Soc (NS)* 43(1):3–23 (electronic)
- Krengel U (1985) *Ergodic theorems*, volume 6 of *de Gruyter studies in mathematics*. Walter de Gruyter, Berlin. With a supplement by Antoine Brunel
- McCutcheon R (1999) *Elemental methods in ergodic Ramsey theory*, volume 1722 of *Lecture notes in mathematics*. Springer, Berlin
- Miles R, Staines M, Ward T (2015) Dynamical invariants for group automorphisms. In: *Recent trends in ergodic theory and dynamical systems*, volume 631 of *Contemp. Math. Amer. Math. Soc.*, Providence, pp 231–258
- Petersen K (1989) *Ergodic theory*, volume 2 of *Cambridge studies in advanced mathematics*. Cambridge University Press, Cambridge. Corrected reprint of the 1983 original
- Schmidt K (1995) *Dynamical systems of algebraic origin*, volume 128 of *Progress in Mathematics*. Birkhäuser, Basel
- Schmidt K (2001) The dynamics of algebraic $\mathbb{Z}^d$ -actions. In: *European congress of mathematics*, Vol. I (Barcelona, 2000), volume 201 of *Progr. Math.* Birkhäuser, Basel, pp 543–553
- Schweiger F (1995) *Ergodic theory of fibred systems and metric number theory*. Oxford Science Publications. The Clarendon Press/Oxford University Press, New York
- Schweiger F (2000) *Multidimensional continued fractions*. Oxford Science Publications. Oxford University Press, Oxford
- Silverman JH (2007) *The arithmetic of dynamical systems*, volume 241 of *Graduate texts in mathematics*. Springer, New York
- Tao T (2006) Arithmetic progressions and the primes. *Collect. Math.*, vol. Extra:37–88
- Tao T (2007a) The dichotomy between structure and randomness, arithmetic progressions, and the primes. In: *International congress of mathematicians*, vol I. European Mathematical Society, Zürich, pp 581–608
- Tao T (2007b) What is good mathematics. *Bull Am Math Soc (NS)* 44(4):623–634
- Totoki H (1969) *Ergodic theory*. *Lecture notes series*, No. 14. Matematisk Institut, Aarhus Universitet, Aarhus
- van der Waerden BL (1971) How the proof of Baude's conjecture was found. In: *Studies in pure mathematics* (presented to Richard Rado). Academic, London, pp 251–260
- Walters P (1982) *An introduction to ergodic theory*, volume 79 of *Graduate texts in mathematics*. Springer, New York
- Weil A (1967) *Basic number theory*. *Die Grundlehren der mathematischen Wissenschaften*, vol 144. Springer, New York
- Ziegler T (2014) Linear equations in primes and dynamics of nilmanifolds. In: *Proceedings of the international congress of mathematicians—Seoul 2014*, vol II. Kyung Moon Sa, Seoul, pp 569–589



Ergodic Theory on Homogeneous Spaces and Metric Number Theory

Dmitry Kleinbock
Department of Mathematics, Brandeis University,
Waltham, MA, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Basic Facts
Further Results
Further Results
Future Directions
Bibliography

Glossary

Diophantine approximation Diophantine approximation refers to approximation of real numbers by rational numbers or, more generally, finding integer points at which some (possibly vector-valued) functions attain values close to integers.

Ergodic theory The study of statistical properties of orbits in abstract models of dynamical systems.

Hausdorff dimension A nonnegative number attached to a metric space and extending the notion of topological dimension of “sufficiently regular” sets, such as smooth submanifolds of real Euclidean spaces.

Homogeneous spaces A homogeneous space G/Γ of a group G by its subgroup Γ is the space of cosets $\{g\Gamma\}$. When G is a Lie group and Γ is a discrete subgroup, the space G/Γ is a smooth manifold and locally looks like G itself.

Lattice; unimodular lattice A lattice in a Lie group is a discrete subgroup of finite covolume; unimodular stands for covolume equal to 1.

Metric number theory Metric number theory (or, specifically, metric Diophantine approximation) refers to the study of sets of real numbers or vectors with prescribed Diophantine approximation properties.

Definition of the Subject

The theory of Diophantine approximation, named after Diophantus of Alexandria, in its simplest set-up deals with the approximation of real numbers by rational numbers. Various higher-dimensional generalizations involve studying values of linear or polynomial maps at integer points. Often a certain “approximation property” is fixed, and one wants to characterize the set of numbers (vectors, matrices) which share this property, by means of certain measures (Lebesgue, or Hausdorff, or some other interesting measures). This is usually referred to as metric Diophantine approximation.

The starting point for the theory is an elementary fact that $\mathbb{Q}$, the set of rational numbers, is dense in $\mathbb{R}$, the reals. In other words, every real number can be approximated by rationals: for any $y \in \mathbb{R}$ and any $\varepsilon > 0$, there exists $p/q \in \mathbb{Q}$ with

$$|y - p/q| < \varepsilon. \quad (1)$$

To answer questions like “how well can various real numbers be approximated by rational numbers? i.e., how small can ε in (1) be chosen for varying $p/q \in \mathbb{Q}$?”, a natural approach has been to compare the accuracy of the approximation of y by p/q to the “complexity” of the latter, which can be measured by the size of its denominator q in its reduced form. This seemingly simple set-up has led to introducing many important Diophantine approximation properties of numbers/vectors/matrices, which show up in various fields of mathematics and physics, such as differential equations, Kolmogorov-Arnold-Moser (KAM) theory, and transcendental number theory.

Introduction

As the first example of refining the statement about the density of $\mathbb{Q}$ in $\mathbb{R}$, consider a theorem by Kronecker stating that for any $y \in \mathbb{R}$ and any $c > 0$, there exist infinitely many $q \in \mathbb{Z}$ such that

$$|y - p/q| < c/|q| \quad \text{i.e. } |qy - p| < c \quad (2)$$

for some $p \in \mathbb{Z}$. A comparison of (1) and (2) shows that it makes sense to multiply both sides of (1) by q , since in the right-hand side of (2) one would still be able to get very small numbers. In other words, approximation of y by p/q translates into approximating integers by integer multiples of y .

Also, if y is irrational, (p, q) can be chosen to be relatively prime, that is, one gets infinitely many different rational numbers p/q satisfying (2). However, if $y \in \mathbb{Q}$, the latter is no longer true for small enough c . Thus, it seems to be more convenient to talk about pairs (p, q) rather than $p/q \in \mathbb{Q}$, avoiding a necessity to consider the two cases separately. At this point, it is convenient to introduce the following central definition: if ψ is a function $\mathbb{N} \rightarrow \mathbb{R}_+$ and $y \in \mathbb{R}$, say that y is ψ -approximable (notation: $y \in \mathcal{W}(\psi)$) if there exist infinitely many $q \in \mathbb{N}$ such that

$$|qy - p| < \psi(q) \quad (3)$$

for some $p \in \mathbb{Z}$. Because of Kronecker's Theorem, it is natural to assume that $\psi(x) \rightarrow 0$ as $x \rightarrow \infty$. Often ψ will be assumed nonincreasing, although many results do not require monotonicity of ψ .

One can similarly consider a higher-dimensional version of the above set-up. Note that $y \in \mathbb{R}$ in the above formulas plays the role of a linear map from $\mathbb{R}$ to another copy of $\mathbb{R}$, and one asks how close values of this map at integers are from integers. It is natural to generalize it by taking a linear operator Y from $\mathbb{R}^n$ to $\mathbb{R}^m$ for fixed $m, n \in \mathbb{N}$, that is, an $m \times n$ -matrix (interpreted as a system of m linear forms Y_i on $\mathbb{R}^n$). We will denote by $M_{m,n}$ the space of $m \times n$ matrices with real coefficients. For ψ as above, one says that $Y \in M_{m,n}$ is ψ -approximable (notation:

$Y \in \mathcal{W}_{m,n}(\psi)$) if there are infinitely many $\mathbf{q} \in \mathbb{Z}^n$ such that

$$\|Y\mathbf{q} + \mathbf{p}\| \leq \psi(\|\mathbf{q}\|) \quad (4)$$

for some $\mathbf{p} \in \mathbb{Z}^m$. Here $\|\cdot\|$ is the supremum norm on $\mathbb{R}^k$ given by $\|\mathbf{y}\| = \max_{1 \leq i \leq k} |y_i|$. (This definition is slightly different from the one used in (Kleinbock and Margulis 1999), which involved powers of norms.)

Traditionally, one of the main goals of metric Diophantine approximation has been to understand how big the sets $\mathcal{W}_{m,n}(\psi)$ are for fixed m, n , and various functions ψ . Of course, (4) is not the only interesting condition that can be studied; various modifications of the approximation properties can also be considered. For example, the Oppenheim Conjecture, now a theorem of Margulis (1989) and a basis for many important developments (Dani and Margulis 1993; Eskin et al. 1998, 2005), states that indefinite irrational quadratic forms can take arbitrary small values at integer points; Littlewood's conjecture, see (18) below, deals with a similar statement about products of linear forms. See the article Ergodic Theory: Rigidity by Nitica and surveys (Einsiedler and Lindenstrauss 2006; Margulis 1997) for details.

We remark that the standard tool for studying Diophantine approximation properties of real numbers ($m = n = 1$) is the continued fraction expansion or, equivalently, the Gauss map $x \mapsto 1/x \bmod 1$ of the unit interval, see (Khinchine 1963). However, the emphasis of this survey lies in higher-dimensional theory, and the dynamical system described below can be thought of as a replacement for the continued fraction technique applicable in the one-dimensional case. Additional details about interactions between ergodic theory and number theory can be found in the article by Nitica mentioned above, in Ergodic Theory: Recurrence by Frantzikinakis and McCutcheon and Ergodic Theory: Interactions with Combinatorics and Number Theory by Ward, as well as in the survey papers (Einsiedler and Lindenstrauss 2006; Eskin 1998; Kleinbock 2001; Kleinbock et al. 2002; Lindenstrauss 2007; Margulis 1997, 2002).

Here is a brief outline of the rest of the article. In the next section, we survey basic results, some classical, some obtained relatively recently, in metric Diophantine approximation. Section “[Connection with Dynamics on the Space of Lattices](#)” is devoted to a description of the connection between Diophantine approximation and dynamics, specifically flows on the space of lattices. In sections “[Diophantine Approximation with Dependent Quantities: The Set-Up](#)” and “[Further Results](#),” we specialize to the set-up of Diophantine approximation on manifolds or, more generally, approximation properties of vectors with respect to measures satisfying some natural conditions and show how applications of homogeneous dynamics contributed to important recent developments in the field. Section “[Future Directions](#)” mentions several open questions and directions for further investigation.

Basic Facts

General references for this section: (Cassels 1957; Schmidt 1980). The starting point for all the explorations in metric number theory is Dirichlet’s Theorem (1842), stating that for any $Y \in M_{m,n}$ and for any $T > 1$, there exist $\mathbf{q} = (q_1, \dots, q_n) \in \mathbb{Z}^n \setminus \{0\}$ and $\mathbf{p} = (p_1, \dots, p_m) \in \mathbb{Z}^m$ satisfying the following system of inequalities:

$$\|Y\mathbf{q} - \mathbf{p}\| < T^{-n/m} \quad \text{and} \quad \|\mathbf{q}\| \leq T. \quad (5)$$

In fact, it is this paper of Dirichlet which gave rise to his box principle. Later another proof of the same result was given by Minkowski.

A natural question to ask is whether one can improve (5) by replacing $T^{-n/m}$ by a smaller function, that is, consider the following system of inequalities:

$$\|Y\mathbf{q} - \mathbf{p}\| < \psi(T) \quad \text{and} \quad \|\mathbf{q}\| \leq T, \quad (5a)$$

where ψ is a positive, continuous, decreasing function which decays to zero at infinity. Historically, there have been two directions to pursue in this regard: looking for solvability of (5a) for an unbounded set of $T > 0$ versus for all large

enough T . The former is sometimes referred to as *asymptotic approximation* and amounts to studying sets $\mathcal{W}_{m,n}(\psi)$ mentioned in the previous section. The latter, less studied set-up of *uniform approximation*, has attracted some attention in recent years. Both set-ups can be approached using tools from dynamics on the space of lattices, as we shall see below.

The simplest choice for functions ψ happens to be the following: let us denote

$$\psi_{c,v}(x) = cx^{-v}.$$

From Dirichlet’s Theorem, it easily follows that $\mathcal{W}_{m,n}(\psi_{1,n/m}) = M_{m,n}$. The constant $c = 1$ is not optimal: the smallest value of c for which $\mathcal{W}_{1,1}(\psi_{c,1}) = \mathbb{R}$ is $1/\sqrt{5}$, and the optimal constants are not known in higher dimensions, although some estimates can be given, see (Schmidt 1980).

Systems of linear forms which do not belong to $\mathcal{W}_{m,n}(\psi_{c,n/m})$ for some positive c are called *badly approximable*; that is, we set

$$\text{BA}_{m,n} \stackrel{\text{def}}{=} M_{m,n} \setminus \bigcup_{c>0} \mathcal{W}_{m,n}(\psi_{c,n/m}).$$

Their existence in arbitrary dimensions was shown by Perron. Note that a real number y ($m = n = 1$) is badly approximable if and only if its continued fraction coefficients are uniformly bounded. It was proved by Jarník (1929) in the case $m = n = 1$ and by Schmidt in the general case (Schmidt 1969) that badly approximable matrices form a set of full Hausdorff dimension: that is, $\dim(\text{BA}_{m,n}) = mn$.

On the other hand, it can be shown that each of the sets $\mathcal{W}_{m,n}(\psi_{c,n/m})$ for any $c > 0$ has full Lebesgue measure, and hence, the complement $\text{BA}_{m,n}$ to their intersection has measure zero. This is a special case of a theorem due to Khintchine (1924) in the case $n = 1$ and to Groshev (1938) in full generality, which gives the precise condition on the function ψ under which the set of ψ -approximable matrices has full measure. Namely, if ψ is nonincreasing (this

assumption can be removed in higher dimensions but not for $n = 1$, see (Duffin and Schaeffer 1941)), then λ -almost no (resp., λ -almost every) $Y \in M_{m,n}$ is ψ -approximable, provided the sum

$$\sum_{k=1}^{\infty} k^{n-1} \psi(k)^m \quad (6)$$

converges (resp., diverges). (Here and hereafter λ stands for Lebesgue measure.) This statement is usually referred to as the Khintchine–Groshev Theorem. The convergence case of this theorem follows in a straightforward manner from the Borel–Cantelli Lemma, but the divergence case is harder. It was reproved and sharpened in 1960 by Schmidt (1960), who showed that if the sum (6) diverges, then for almost all Y , the number of solutions to (4) with $\|\mathbf{q}\| \leq T$ is asymptotic to the partial sum of the series (6) (up to a constant) and also gave an estimate for the error term. A special case of the convergence part of the theorem shows that $\mathcal{W}_{m,n}(\psi_{1,v})$ has measure zero whenever $v > n/m$. Y is said to be *very well approximable* if it belongs to $\mathcal{W}_{m,n}(\psi_{1,v})$ for some $v > n/m$. That is,

$$\text{VWA}_{m,n} \stackrel{\text{def}}{=} \cup_{v > n/m} \mathcal{W}_{m,n}(\psi_{1,v}).$$

More specifically, let us define the *Diophantine exponent* $\omega(Y)$ of Y (sometimes called “the exact order” of Y) to be the supremum of $v > 0$ for which $Y \in \mathcal{W}_{m,n}(\psi_{1,v})$. Then $\omega(Y)$ is always not less than n/m and is equal to n/m for Lebesgue-a.e. Y ; in fact, $\text{VWA}_{m,n} = \{Y \in M_{m,n} : \omega(Y) > n/m\}$. The Hausdorff dimension of the null sets $\mathcal{W}_{m,n}(\psi_{1,v})$ was computed independently by Besicovitch (1929) and Jarník (1928) in the one-dimensional case and by Dodson (1992) in general: when $v > n/m$, one has

$$\dim(\mathcal{W}_{m,n}(\psi_{1,v})) = (n-1)m + \frac{m+n}{v+1}. \quad (7)$$

See (Dodson 1993) for a nice exposition of ideas involved in the proof of both the aforementioned formula and the Khintchine–Groshev Theorem. Note that it follows from (7) that the null set

$\text{VWA}_{m,n}$ has full Hausdorff dimension. Matrices contained in the intersection

$$\bigcap_{v>0} \mathcal{W}_{m,n}(\psi_{1,v}) = \{Y \in M_{m,n} : \omega(Y) = \infty\}$$

are called *Liouville* and form a set of Hausdorff dimension $(n-1)m$, that is, to the dimension of Y for which $Y\mathbf{q} \in \mathbb{Z}$ for some $\mathbf{q} \in \mathbb{Z}^n \setminus \{0\}$ (the latter belongs to $\mathcal{W}_{m,n}(\psi)$ for any positive ψ).

Metric theory of uniform approximation was initiated by Davenport and Schmidt in (1969/1970, 1970). Let us say that $Y \in M_{m,n}$ is *ψ -Dirichlet* (notation: $Y \in \mathcal{D}_{m,n}(\psi)$) if the system of inequalities (5a) has solutions in $(\mathbf{p}, \mathbf{q}) \in \mathbb{Z}^m \times (\mathbb{Z}^n \setminus \{0\})$ for all sufficiently large T . (Again, this definition is slightly different from the one used in several papers on the subject (Kleinbock and Wadleigh 2018, 2019; Kleinbock et al. 2021a), where powers of norms were considered.) Dirichlet’s Theorem asserts that $\mathcal{D}_{m,n}(\psi_{1,n/m}) = M_{m,n}$. However, the constant 1 is now optimal: it was shown in (Davenport and Schmidt 1970) that the Lebesgue measure of $\mathcal{D}_{m,n}(\psi_{c,n/m})$ is zero for any $c < 1$. This phenomenon can be easily explained by means of dynamics on homogeneous spaces, see (Dani 1985; Kleinbock and Weiss 2008). Y is said to be *Dirichlet-Improvable* if it belongs to the null set $\text{DI}_{m,n} \stackrel{\text{def}}{=} \cup_{c<1} \mathcal{D}_{m,n}(\psi_{c,n/m})$, and *singular* if it belongs to $\text{SING}_{m,n} \stackrel{\text{def}}{=} \cap_{c>0} \mathcal{D}_{m,n}(\psi_{c,n/m})$. Note that Schmidt also showed that $\text{DI}_{m,n}$ contains $\text{BA}_{m,n}$, hence is a set of full Hausdorff dimension.

Note also that all the aforementioned properties behave nicely with respect to transposition; this is described by the so-called Khintchine’s Transference Principle (Chapter V in (Cassels 1957)). For example, $Y \in \text{BA}_{n,m}$ if and only if $Y^T \in \text{BA}_{n,m}$, and $Y \in \text{VWA}_{m,n}$ if and only if $Y^T \in \text{VWA}_{n,m}$. In particular, many problems related to approximation properties of vectors ($n = 1$) and linear forms ($m = 1$) reduce to one another.

We refer the readers to (Beresnevich et al. 2006; Harman 1998) for very detailed and comprehensive accounts of various further aspects of the theory.

Connection with Dynamics on the Space of Lattices

General references for this section: (Bekka and Mayer 2000; Starkov 2000).

Interactions between Diophantine approximation and the theory of dynamical systems has a long history. Already in Kronecker's Theorem one can see a connection. Indeed, the statement of the theorem can be rephrased as follows: the points on the orbit of 0 under the rotation of the circle $\mathbb{R}/\mathbb{Z}$ by y approach the initial point 0 arbitrarily closely. This is a special case of the Poincaré Recurrence Theorem in measurable dynamics. And, likewise, all the aforementioned properties of $Y \in M_{m,n}$ can be restated in terms of recurrence properties of the $\mathbb{Z}^n$ -action on the m -dimensional torus $\mathbb{R}^m/\mathbb{Z}^m$ given by $\mathbf{x} \mapsto Y\mathbf{x} \bmod \mathbb{Z}^m$. In other words, fixing Y gives rise to a dynamical system in which approximation properties of Y show up.

However, the theme of this section is a different dynamical system, whose phase space is (essentially) the space of parameters Y and which can be used to read the properties of Y from the behavior of the associated trajectory.

It has been known for a long time (see (Sheingorn 1993) for a historical account) that Diophantine properties of real numbers can be coded by the behavior of geodesics on the quotient of the hyperbolic plane by $\mathrm{SL}_2(\mathbb{Z})$. In fact, the latter flow can be viewed as the suspension flow of the Gauss map mentioned at the end of section “Introduction.” There have been many attempts to construct a higher-dimensional analogue of the Gauss map so that it captures all the features of simultaneous approximation, see (Khanin et al. 2007; Kontsevich and Suhov 1999; Lagarias 1994) and references therein. On the other hand, it seems to be more natural and efficient to generalize the suspension flow itself, and this is where one needs higher rank homogeneous dynamics.

As was mentioned above, in the basic set-up of simultaneous Diophantine approximation, one takes a system of m linear forms $Y_1, \dots, Y_m$ on $\mathbb{R}^n$ and looks at the values of $|Y_i(\mathbf{q}) + p_i|$, $p_i \in \mathbb{Z}$, when $\mathbf{q} = (q_1, \dots, q_n) \in \mathbb{Z}^n$ is far from 0. The trick is to put together

$$Y_1(\mathbf{q}) + p_1, \dots, Y_m(\mathbf{q}) + p_m \quad \text{and} \quad q_1, \dots, q_n,$$

and consider the collection of vectors

$$\left\{ \begin{pmatrix} Y\mathbf{q} + \mathbf{p} \\ \mathbf{q} \end{pmatrix} \middle| \mathbf{p} \in \mathbb{Z}^m, \mathbf{q} \in \mathbb{Z}^n \right\} = L_Y \mathbb{Z}^k$$

where $k = m + n$ and

$$L_Y \stackrel{\text{def}}{=} \begin{pmatrix} I_m & Y \\ 0 & I_n \end{pmatrix}, \quad Y \in M_{m,n}. \quad (8)$$

This collection is a unimodular lattice in $\mathbb{R}^k$, that is, a discrete subgroup of $\mathbb{R}^k$ with covolume 1. Our goal is to keep track of vectors in such a lattice having small projections onto the first m components of $\mathbb{R}^k$ and big projections onto the last n components. This is where dynamics comes into the picture. Denote by g_t the one-parameter subgroup of $\mathrm{SL}_k(\mathbb{R})$ given by

$$g_t = \text{diag} \left(\underbrace{e^{t/m}, \dots, e^{t/m}}_{m \text{ times}}, \underbrace{e^{-t/n}, \dots, e^{-t/n}}_{n \text{ times}} \right). \quad (9)$$

The vectors in the lattice $L_Y \mathbb{Z}^k$ are moved by the action of g_t , $t > 0$, and a special role is played by the moment t when the “small” and “big” projections equalize. That is, one is led to consider a new dynamical system. Its phase space is the space of unimodular lattices in $\mathbb{R}^k$, which can be naturally identified with the homogeneous space

$$\Omega_k \stackrel{\text{def}}{=} G/\Gamma, \quad \text{where} \quad (10)$$

$$G = \mathrm{SL}_k(\mathbb{R}) \text{ and } \Gamma = \mathrm{SL}_k(\mathbb{Z}),$$

and the action is given by left multiplication by elements of the subgroup (9) of G or perhaps other subgroups $H \subset G$. Study of such systems has a rich history; for example, they are known to be ergodic and mixing whenever H is unbounded (Moore 1966). What is important in this particular case is that the space Ω_k happens to be non-compact, and its structure at infinity is described via Mahler's Compactness Criterion, see Chapter V in (Bekka and Mayer 2000): a sequence

of lattices $g_i \mathbb{Z}^k$ goes to infinity in $\Omega_k \Leftrightarrow$ there exists a sequence $\{\mathbf{v}_i \in \mathbb{Z}^k \setminus \{0\}\}$ such that $g_i(\mathbf{v}_i) \rightarrow 0$ as $i \rightarrow \infty$. Equivalently, for $\varepsilon > 0$ consider a subset K_ε of Ω_k consisting of lattices with no nonzero vectors of norm less than ε ; then all the sets K_ε are compact, and every compact subset of Ω_k is contained in one of them. Moreover, one can choose a metric on Ω_k such that $\text{dist}(\Lambda, \mathbb{Z}^k)$ is, up to a uniform multiplicative constant, equal to $-\log \min_{\mathbf{v} \in \Lambda \setminus \{0\}} \|\mathbf{v}\|$ (see (Ding 1994)), then the length of the smallest nonzero vector in a lattice Λ will determine how far away is this lattice in the “cusp” of Ω_k . Note that Minkowski’s Convex Body Theorem shows that $K_\varepsilon = \emptyset$ when $\varepsilon > 1$. On the other hand, when $\varepsilon < 1$, each of the sets K_ε has nonempty interior, and $K_1 = \bigcap_{\varepsilon < 1} K_\varepsilon$ is nonempty, but has empty interior and zero Haar measure. It is called the *critical locus* for the supremum norm, and its structure is described by the Hajós–Minkowski Theorem.

Using Mahler’s Criterion, it is not hard to show that $Y \in \text{BA}_{m,n}$ if and only if the trajectory

$$\{g_t L_Y \mathbb{Z}^k : t \in \mathbb{R}_+\} \quad (11)$$

is bounded in Ω_k . This was proved by Dani (1985) in 1985 and later generalized in (Kleinbock and Margulis 1999) to produce a criterion for Y to be ψ -approximable for any nonincreasing function ψ . Namely, given ψ one can in a unique way define the function $r(t)$ such that $Y \in \mathcal{W}_{m,n}(\psi)$ if and only if $g_t L_Y \mathbb{Z}^k \notin K_{r(t)}$ for an unbounded set of $t > 0$. An important special case is a criterion for a system of linear forms to be very well approximable: $Y \in \text{VWA}_{m,n}$ if and only if the trajectory (11) has linear growth, that is, there exists a positive γ such that $\text{dist}(g_t L_Y \mathbb{Z}^k, \mathbb{Z}^k) > \gamma t$ for an unbounded set of $t > 0$. Similarly, one can produce a uniform version, that is, a criterion for Y to be ψ -Dirichlet: $Y \in \mathcal{D}_{m,n}(\psi)$ if and only if $g_t L_Y \mathbb{Z}^k \notin K_{r(t)}$ for all large enough $t > 0$. In particular, Y is singular if and only if $\{g_t L_Y \mathbb{Z}^k\}$ is divergent and is Dirichlet-Improvable if and only if $g_t L_Y \mathbb{Z}^k$ is not in K_ε for some $\varepsilon > 0$; that is, the g_t trajectory of $L_Y \mathbb{Z}^k$ eventually stays away from the critical locus K_1 .

This correspondence allows one to link various Diophantine and dynamical phenomena. For example, from the results of (Kleinbock and Margulis 1996) on abundance of bounded orbits on homogeneous spaces, one can deduce the aforementioned theorem of Schmidt (1969): the set $\text{BA}_{m,n}$ has full Hausdorff dimension. See also Kleinbock and Weiss (2013) where a winning property of $\text{BA}_{m,n}$ is established by dynamical methods. A dynamical Borel–Cantelli Lemma established in (Kleinbock and Margulis 1999) can be used for an alternative proof of the Khintchine–Groshev Theorem; see also Sullivan (1982) for an earlier geometric approach. Note that establishing a uniform approximation analogue of the Khintchine–Groshev Theorem, that is, finding a criterion for sets $\mathcal{D}_{m,n}(\psi)$ to have zero or full measure, is still an open problem; it was solved in (Kleinbock and Wadleigh 2018) for $m = n = 1$ using continued fractions, and a partial solution for the general case was developed in Kleinbock et al. (2021a). It is worthwhile to point out that all the aforementioned proofs are based on the following two properties of the g_t -action: effective mixing, which forces points to return to compact subsets and makes preimages of cusp neighborhoods quasi-independent, and hyperbolicity, which implies that the behavior of points on unstable leaves is generic. The latter is important since the orbits of the group $\{L_Y : Y \in M_{m,n}\}$ are precisely the unstable leaves with respect to the g_t -action.

$$\|Y\mathbf{q} - \mathbf{p}\| < \psi(T) \quad \text{and} \quad \|\mathbf{q}\| \leq T, \quad (11a)$$

We note that other types of Diophantine problems, such as conjectures of Oppenheim and Littlewood mentioned in the previous section, can be reduced to statements involving actions on Ω_k by means of the same principle: Mahler’s Criterion is used to relate small values of some function at integer points to excursions to infinity in Ω_k of orbit of the stabilizer of this function. In particular, Littlewood’s Conjecture deals with *multiplicative approximation* and is related to a multi-parameter action by a certain semigroup of diagonal matrices, see the section “Further

Results” part of section “**Diophantine Approximation with Dependent Quantities: The Set-Up.**”

Other important and useful recent applications of homogeneous dynamics to metric Diophantine approximation are related to the circle of ideas roughly called “Diophantine approximation with dependent quantities” (terminology borrowed from (Sprindžuk 1979)), to be surveyed in the next two sections.

Further Results

Overall there have been numerous developments utilizing the dynamical approach to Diophantine problems. For example, the Hausdorff dimension of the null set $\text{SING}_{m,n}$ of singular systems of linear forms was computed for $m = 2, n = 1$ by Cheung (2011) and then for $n = 1$ and arbitrary m by Cheung and Chevallier (2016). In the general case, the sharp upper estimate was obtained in (Kadyrov et al. 2017), where dynamics was used via the method of integral inequalities for height functions on the space of lattices originally developed in (Eskin et al. 1998); those are also known as *Margulis functions*, see (Eskin and Mozes 2022) for a recent survey. A refinement of the technique from (Kadyrov et al. 2017), combined with the exponential mixing of the g_t -action on Ω_k , produces a proof of the Dimension Drop Conjecture (Kleinbock and Mirzadeh 2020); its byproduct is a sharpening of the Davenport–Schmidt result, namely, establishing that $\dim(\mathcal{D}_{m,n}(\psi_{c,n/m})) < mn$ for any $c < 1$.

The complimentary lower estimate for the dimension of the set of singular matrices, producing the equality

$$\dim(\text{SING}_{m,n}) = mn - \frac{mn}{m+n},$$

was established in (Das et al. 2019) using the method of *parametric geometry of numbers*, stemming from the work of Schmidt–Summerer (2013) and Roy (2015). The latter method is a powerful tool that gives a strong quantitative way to relate Diophantine properties of matrices Y to the

behavior of the trajectory $\{g_t L_Y \mathbb{Z}^k\}$. One of its applications is a better understanding of *uniform Diophantine exponents* of matrices; another one is showing that the set $\text{DI}_{m,n} \setminus (\text{BA}_{m,n} \cup \text{SING}_{m,n})$ is uncountable, see (Beresnevich et al. 2020).

In another development, the correspondence between approximation and dynamics can be extended to inhomogeneous approximation, that is, studying values $Y\mathbf{q} + \mathbf{z} + \mathbf{p}$ of *systems of affine forms* (here $Y \in M_{m,n}$ and $\mathbf{z} \in \mathbb{R}^m$) at integer points $(\mathbf{q}, \mathbf{p})$. This way one obtains a zero-infinity law for improvement of Dirichlet’s theorem, see (Kleinbock and Wadleigh 2019) and also (Kim and Kim 2022) for an alternative proof and a version for Hausdorff measures.

For recent results utilizing the correspondence between Diophantine approximation and homogeneous dynamics, see (Athreya et al. 2015; Dolgopyat et al. 2017; Björklund and Gorodnik 2019; Shapira and Weiss 2022) among many other papers.

Diophantine Approximation with Dependent Quantities: The Set-Up

General references for this section: (Bernik and Dodson 1999; Sprindžuk 1979).

Here we restrict ourselves to Diophantine properties of vectors in $\mathbb{R}^n$. In particular, we will look more closely at the set of very well approximable vectors, which we will simply denote by VWA , dropping the subscripts. In many cases, it does not matter whether one works with row or column vectors, in view of the duality remark made at the end of section “**Basic Facts.**” Note however that during recent years there has been a substantial progress toward establishing analogues of results from this section to the space of matrices (this was mentioned as work in progress in [Kleinbock and Margulis 1998]), see (Kleinbock et al. 2010; Beresnevich et al. 2015; Aka et al. 2018; Das et al. 2018; Yang 2019).

We begin with a non-example of an application of dynamics to Diophantine approximation: a celebrated and difficult theorem which currently, to the best of the author’s knowledge, has no dynamical proof. Suppose that $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$ is such that each y_i is algebraic and $1, y_1, \dots, y_n$ are

linearly independent over $\mathbb{Q}$. It was established by Roth for $n = 1$ (Roth 1955) and then generalized to arbitrary n by Schmidt (1972) that y as above necessarily belongs to the complement of VWA. In other words, vectors with very special algebraic properties happen to follow the behavior of a generic vector in $\mathbb{R}^n$.

We would like to view the above example as a special case of a general class of problems. Namely, suppose we are given a Radon measure μ on $\mathbb{R}^n$. Let us say that μ is extremal (Sprindžuk 1980) if μ -a.e. $y \in \mathbb{R}^n$ is not very well approximable. Further, define the *Diophantine exponent* $\omega(\mu)$ of μ to be the μ -essential supremum of the function $\omega(\cdot)$; in other words,

$$\omega(\mu) \stackrel{\text{def}}{=} \sup \{v | \mu(\mathcal{W}(\psi_{1,v})) > 0\}.$$

Clearly it only depends on the measure class of μ . If μ is naturally associated with a subset $\mathcal{M}$ of $\mathbb{R}^n$ supporting μ (e.g., if $\mathcal{M}$ is a smooth submanifold of $\mathbb{R}^n$ and μ is the measure class of the Riemannian volume on $\mathcal{M}$, or, equivalently, the pushforward $\mathbf{f}_*\lambda$ of λ by a smooth map f parametrizing $\mathcal{M}$), one defines the Diophantine exponent $\omega(\mathcal{M})$ of $\mathcal{M}$ to be equal to that of μ and says that $\mathcal{M}$ is *extremal* if $\mathbf{f}(\mathbf{x})$ is not very well approximable for μ -a.e. $\mathbf{x}$.

Then $\omega(\mu) \geq n$ for any μ , and $\omega(\lambda) = \omega(\mathbb{R}^n)$ is equal to n . The latter justifies the use of the word “extremal”: μ is extremal if $\omega(\mu)$ is equal to n , that is, attains the smallest possible value. The aforementioned results of Roth and Schmidt then can be interpreted as the extremality of atomic measures supported on algebraic vectors without rational dependence relations.

Historically, the first measure (other than λ) to be considered in the set-up described above was the pushforward of λ by the map

$$\mathbf{f}(x) = (x, x^2, \dots, x^n). \quad (12)$$

The extremality of $\mathbf{f}_*\lambda$ for f as above was conjectured in 1932 by K. Mahler (1932) and proved in 1964 by Sprindžuk (1964, 1969). It was important for Mahler’s study of transcendental numbers: this result, roughly speaking, says that almost all transcendental numbers are “not

very algebraic.” At about the same time, Schmidt (1964) proved the extremality of $\mathbf{f}_*\lambda$ when $\mathbf{f}: I \rightarrow \mathbb{R}^2$, $I \subset \mathbb{R}$, is C^3 and satisfies

$$\begin{vmatrix} f'_1(x) & f'_2(x) \\ f''_1(x) & f''_2(x) \end{vmatrix} \neq 0 \quad \text{for } \lambda - \text{a.e. } x \in I;$$

in other words, the curve parametrized by f has nonzero curvature at almost all points. Since then, a lot of attention has been devoted to showing that measures $\mathbf{f}_*\lambda$ are extremal for other smooth maps f .

To describe a broader class of examples, recall the following definition. Let $\mathbf{x} \in \mathbb{R}^d$ and let $\mathbf{f} = (f_1, \dots, f_n)$ be a C^k map from a neighborhood of $\mathbf{x}$ to $\mathbb{R}^n$. Say that f is *nondegenerate* at $\mathbf{x}$ if $\mathbb{R}^n$ is spanned by partial derivatives of f at $\mathbf{x}$ up to some order. Say that f is *nondegenerate* if it is nondegenerate at λ -a.e. $\mathbf{x}$. It was conjectured by Sprindžuk (1979) in 1980 that $\mathbf{f}_*\lambda$ for real analytic nondegenerate f are extremal. Many special cases were established since then (see (Bernik and Dodson 1999) for a detailed exposition of the theory and many related results), but the general case stood open until the mid-1990s (Kleinbock and Margulis 1998), when Sprindžuk’s conjecture was proved using the dynamical approach (later Beresnevich (2002) succeeded in establishing and extending this result without use of dynamics). The proof in (Kleinbock and Margulis 1998) uses the correspondence outlined in the previous section plus a measure estimate for flows on the space of lattices which is described below.

In the subsequent work, the method of (Kleinbock and Margulis 1998) was adapted to a much broader class of measures. To define it we need to introduce some more notation and definitions. If $\mathbf{x} \in \mathbb{R}^d$ and $r > 0$, denote by $B(\mathbf{x}, r)$ the open ball of radius r centered at $\mathbf{x}$. If $B = B(\mathbf{x}, r)$ and $c > 0$, cB will denote the ball $B(\mathbf{x}, cr)$. For $B \subset \mathbb{R}^d$ and a real-valued function f on B , let

$$\|f\|_B \stackrel{\text{def}}{=} \sup_{\mathbf{x} \in B} |f(\mathbf{x})|.$$

If ν is a measure on $\mathbb{R}^d$ such that $\nu(B) > 0$, define $\|f\|_{\nu, B} \stackrel{\text{def}}{=} \|f\|_{B \cap \text{supp } \nu}$; this is the same as the $L^\infty(\nu)$ -norm of $f|_B$ if f is continuous and B is

open. If $D > 0$ and $U \subset \mathbb{R}^d$ is an open subset, let us say that ν is D -Federer on U if for any ball $B \subset U$ centered at $\text{supp } \nu$, one has $\frac{\nu(3B)}{\nu(B)} < D$ whenever $3B \subset U$. This condition is often called “doubling” in the literature. See (Kleinbock et al. 2004; Mauldin and Urbański 1996) for examples and references. ν is called Federer if for ν -a.e. $\mathbf{x} \in \mathbb{R}^d$, there exist a neighborhood U of $\mathbf{x}$ and $D > 0$ such that ν is D -Federer on U .

Given $C, \alpha > 0$, open $U \subset \mathbb{R}^d$, and a measure ν on U , a function $f: U \rightarrow \mathbb{R}$ is called (C, α) -good on U with respect to ν if for any ball $B \subset U$ centered in $\text{supp } \nu$ and any $\varepsilon > 0$, one has

$$\begin{aligned} \nu(\{\mathbf{x} \in B : |f(\mathbf{x})| < \varepsilon\}) \\ \leq C \left(\frac{\varepsilon}{\|f\|_{\nu, B}} \right)^\alpha \nu(B). \end{aligned} \quad (13)$$

This condition was formally introduced in (Kleinbock and Margulis 1998) for ν being Lebesgue measure and in (Kleinbock et al. 2004) for arbitrary ν . A basic example is given by polynomials, and the upshot of the above definition is the formalization of a property needed for the proof of several basic facts (Dani 1979, 1986; Margulis 1975) about polynomial maps into the space of lattices. In (Kleinbock et al. 2004) a strengthening of this property was considered: f was called absolutely (C, α) -good on U with respect to ν if for B and ε as above one has

$$\nu(\{\mathbf{x} \in B : |f(\mathbf{x})| < \varepsilon\}) \leq C \left(\frac{\varepsilon}{\|f\|_B} \right)^\alpha \nu(B). \quad (14)$$

There is no difference between (13) and (14) when ν has full support, but it turns out to be useful for describing measures supported on proper (e.g., fractal) subsets of $\mathbb{R}^d$.

Now suppose that we are given a measure ν on $\mathbb{R}^d$, an open $U \subset \mathbb{R}^d$ with $\nu(U) > 0$, and a map $\mathbf{f} = (f_1, \dots, f_n): \mathbb{R}^d \rightarrow \mathbb{R}^n$. Following (Kleinbock and Weiss 2008), say that a pair $(\mathbf{f}, \nu)$ is (absolutely) good on U if any linear combination of $1, f_1, \dots, f_n$ is (absolutely) (C, α) -good on U with respect to ν . If for ν -a.e. $\mathbf{X}$, there exists a neighborhood U of $\mathbf{X}$ and $C, \alpha > 0$ such that ν is (absolutely) (C, α) -good on U , we will say that the pair $(\mathbf{f}, \nu)$ is (absolutely) good.

Another relevant notion is the nonplanarity of $(\mathbf{f}, \nu)$. Namely, $(\mathbf{f}, \nu)$ is said to be *nonplanar* if whenever B is a ball with $\nu(B) > 0$, the restrictions of $1, f_1, \dots, f_n$ to $B \cap \text{supp } \nu$ are linearly independent over $\mathbb{R}$; in other words, $\mathbf{f}(B \cap \text{supp } \nu)$ is not contained in any proper affine subspace of $\mathbb{R}^n$. Note that absolutely good implies both good and nonplanar, but the converse is in general not true.

Many examples of (absolutely) good and nonplanar pairs $(\mathbf{f}, \nu)$ can be found in the literature. Already the case $n = d$ and $\mathbf{f} = \text{Id}$ is very interesting. A measure μ on $\mathbb{R}^n$ is said to be *friendly* (respectively, *absolutely friendly*) if and only if it is Federer and the pair (Id, μ) is good and nonplanar (respectively, absolutely good). See (Kleinbock et al. 2004; Stratmann and Urbanski 2006; Urbanski 2005) for many examples. An important class of measures is given by limit measures of irreducible system of self-similar or self-conformal contractions satisfying the Open Set Condition (Hutchinson 1981); those are shown to be absolutely friendly in (Kleinbock et al. 2004). The prime example is the middle-third Cantor set on the real line. The term “friendly” was cooked up as a loose abbreviation for “Federer, nonplanar and decaying” and later proved to be particularly friendly in dealing with problems arising in metric number theory, see, for example, (Fishman 2006).

Also let us say that a pair $(\mathbf{f}, \nu)$ is *non-degenerate* if f is nondegenerate at ν -a.e. $\mathbf{x}$. When ν is Lebesgue measure on $\mathbb{R}^d$, it is proved in Proposition 3.4 in (Kleinbock and Margulis 1998) that a nondegenerate $(\mathbf{f}, \nu)$ is good and nonplanar. The same conclusion is derived in Proposition 7.3 in (Kleinbock et al. 2004), assuming that ν is absolutely friendly. Thus, volume measures on smooth nondegenerate manifolds are friendly, but not absolutely friendly.

It turns out that all the aforementioned examples of measures can be proved to be extremal by a generalization of the argument from (Kleinbock and Margulis 1998). Specifically, let ν be a Federer measure on $\mathbb{R}^d$, U an open subset of $\mathbb{R}^d$, and $\mathbf{f}: U \rightarrow \mathbb{R}^n$ a continuous map such that the pair $(\mathbf{f}, \nu)$ is good and nonplanar, then $\mathbf{f}_* \nu$ is extremal.

This can be derived from the Borel–Cantelli Lemma, the correspondence described in the previous section, and the following measure estimate: if ν , U , and f are as above, then for ν -a.e. $\mathbf{x}_0 \in U$, there exists a ball $B \subset U$ centered at $\mathbf{x}_0$ and $\tilde{C}, \alpha > 0$ such that for any $t \in \mathbb{R}_+$ and any $\varepsilon > 0$,

$$\nu(\{\mathbf{x} \in B : g_t L_{\mathbf{f}(\mathbf{x})} \mathbb{Z}^{n+1} \notin K_\varepsilon\}) < \tilde{C} \varepsilon^\alpha. \quad (15)$$

Here g_t is as in (9) with $m = 1$ (assuming that the row vector viewpoint is adopted). This is a quantitative way of saying that for fixed t , the “flow” $\mathbf{x} \mapsto g_t L_{\mathbf{f}(\mathbf{x})} \mathbb{Z}^{n+1}$, $B \rightarrow \Omega_{n+1}$ cannot diverge and in fact must spend a big (uniformly in t) proportion of time inside compact sets K_ε .

The inequality (15) is derived from a general “quantitative non-divergence” estimate, which can be thought of a substantial generalization of theorems of Margulis and Dani (Dani 1979, 1986; Margulis 1975) on nondivergence of unipotent flows on homogeneous spaces. One of its most general versions (Kleinbock et al. 2004) deals with a measure ν on $\mathbb{R}^d$, a continuous map $h : \tilde{B} \rightarrow G$, where $\tilde{B}$ is a ball in $\mathbb{R}^d$ centered at $\text{supp } \nu$ and G is as in (10). To describe the assumptions on h , one needs to employ the combinatorial structure of lattices in $\mathbb{R}^k$, and it will be convenient to use the following notation: if V is a nonzero rational subspace of $\mathbb{R}^k$ and $g \in G$, define $\ell_V(g)$ to be the covolume of $g(V \cap \mathbb{Z}^k)$ in gV . Then, given positive constants C, D, α , there exists $C_1 = C_1(d, k, C, \alpha, D) > 0$ with the following property. Suppose ν is D -Federer on $\tilde{B}$, $0 < \rho \leq 1$, and h is such that for each rational $V \subset \mathbb{R}^k$, $\ell_V \circ h$ is (C, α) -good on $\tilde{B}$ with respect to ν , and $\|\ell_V \circ h\|_{\nu, B} \geq \rho$, where $B = 3^{-(k-1)} \tilde{B}$. Then for any positive $\varepsilon \leq \rho$, one has

$$\begin{aligned} \nu(\{\mathbf{x} \in B : h(\mathbf{x}) \mathbb{Z}^k \notin K_\varepsilon\}) \\ \leq C_1 (\varepsilon \rho)^\alpha \nu(B). \end{aligned} \quad (16)$$

Taking $h(\mathbf{x}) = g_t L_{\mathbf{f}(\mathbf{x})}$ and unwinding the definitions of good and nonplanar pairs, one can show that (i) and (ii) can be verified for some balls B centered at ν -almost every point, and derive (15)

from (16). For more detail on the proof, see the lecture notes (Kleinbock 2010).

Further Results

The approach to metric Diophantine approximation using quantitative nondivergence, that is, the implication (i) + (ii) $\Rightarrow$ (iii), is not omnipotent. In particular, it is difficult to use when more precise results are needed, such as computing/estimating the Hausdorff dimension of the set of $\psi_{1,\nu}$ -approximable vectors on a manifold. See (Beresnevich et al. 2007; Beresnevich and Velani 2007) for such results. On the other hand, the dynamical approach can often treat much more general objects than its classical counterpart, and also can be perturbed in a lot of directions, producing many generalizations and modifications of the main theorems from the preceding section. It has numerous applications, most of which show up in a recent survey (Beresnevich and Kleinbock 2022).

One of the most important of them is the so-called multiplicative version of the set-up of section “[Diophantine Approximation with Dependent Quantities: The Set-Up](#).” Namely, define functions $\Pi(\mathbf{x}) \stackrel{\text{def}}{=} \prod_i |x_i|$ and $\Pi_+(\mathbf{x}) \stackrel{\text{def}}{=} \prod_i \max(|x_i|, 1)$. Then, given a function $\psi : \mathbb{N} \rightarrow \mathbb{R}_+$, one says that $Y \in M_{m,n}$ is *multiplicatively ψ -approximable* (notation: $Y \in \mathcal{W}_{m,n}^\times(\psi)$) if there are infinitely many $\mathbf{q} \in \mathbb{Z}^n$ such that

$$\Pi(Y\mathbf{q} + \mathbf{p})^{1/m} \leq \psi(\Pi_+(\mathbf{q})^{1/n}) \quad (17)$$

for some $\mathbf{p} \in \mathbb{Z}^m$. Since $\Pi(\mathbf{x}) \leq \Pi_+(\mathbf{x}) \leq \|\mathbf{x}\|^k$ for $\mathbf{x} \in \mathbb{R}^k$, any ψ -approximable Y is multiplicatively ψ -approximable; but the converse is in general not true, see, for example, (Gallagher 1962). However if one, as before, considers the family $\{\psi_{1,\nu}\}$, the critical parameter for which the drop from full measure to measure zero occurs is again n/m . That is, if one defines the multiplicative Diophantine exponent $\omega^\times(Y)$ of Y by $\omega^\times(Y) \stackrel{\text{def}}{=} \sup\left\{\nu : Y \in \mathcal{W}_{m,n}^\times(\psi_{1,\nu})\right\}$, then clearly $\omega^\times(Y) \geq \omega(Y)$ for all Y , and yet $\omega^\times(Y) = n/m$ for λ -a.e. $Y \in M_{m,n}$.

Now specialize to $\mathbb{R}^n$ (by the same duality principle as before, it does not matter whether to think in terms of row or column vectors, but we will adopt the row vector set-up), and define the *multiplicative exponent* $\omega^\times(\mu)$ of a measure μ on $\mathbb{R}^n$ by $\omega^\times(\mu) \stackrel{\text{def}}{=} \sup\{v | \mu(\mathcal{W}^\times(\psi_{1,v})) > 0\}$; then $\omega^\times(\lambda) = n$. Following Sprindžuk (1980), say that μ is strongly extremal if $\omega^\times(\mu) = n$. It turns out that all the results mentioned in the previous section have their multiplicative analogues; that is, the measures described there happen to be strongly extremal. This was conjectured by A. Baker (1975) for the curve (12) and then by Sprindžuk in 1980 (Sprindžuk 1980) for analytic non-degenerate manifolds. (We remark that only very few results in this set-up can be obtained by the standard methods, see, e.g., (Beresnevich and Velani 2007)). The proof of this stronger statement is based on using the multi-parameter action of

$$g_{\mathbf{t}} = \text{diag}(e^{t_1+\dots+t_n}, e^{-t_1}, \dots, e^{-t_n}), \text{ where} \\ \mathbf{t} = (t_1, \dots, t_n)$$

instead of g_t considered in the previous section. One can show that the choice $h(\mathbf{x}) = g_{\mathbf{t}}L_{\mathbf{t}(\mathbf{x})}$ allows one to verify (i) and (ii) uniformly in $\mathbf{t} \in \mathbb{R}_+^n$, and the proof is finished by applying a multi-parameter version of the correspondence described in section “[Connection with Dynamics on the Space of Lattices](#).” Namely, one can show that $\mathbf{y} \in \text{VWA}_{1,n}^\times$ if and only if the trajectory $\{g_{\mathbf{t}}L_{\mathbf{y}}\mathbb{Z}^k : \mathbf{t} \in \mathbb{R}_+^n\}$ grows linearly, that is, for some $\gamma > 0$, one has $\text{dist}(g_{\mathbf{t}}L_{\mathbf{y}}\mathbb{Z}^{n+1}, \mathbb{Z}^{n+1}) > \gamma \|\mathbf{t}\|$ for an unbounded set of $\mathbf{t} \in \mathbb{R}_+^n$. A similar correspondence was used in (Einsiedler et al. 2006) to prove that the (conjecturally empty) set of exceptions to Littlewood’s Conjecture, which, using the terminology introduced above, can be called badly multiplicatively approximable vectors:

$$\text{BA}_{n,1}^\times \stackrel{\text{def}}{=} \mathbb{R}^n \setminus \bigcup_{c>0} \mathcal{W}_{n,1}^\times(\psi_{c,1/n}) \\ = \left\{ \mathbf{y} : \inf_{q \in \mathbb{Z} \setminus \{0\}, \mathbf{p} \in \mathbb{Z}^n} |q| \cdot \Pi(q\mathbf{y} - \mathbf{p}) > 0 \right\}, \quad (18)$$

has Hausdorff dimension zero. This was done using a measure rigidity result for the action of

the group of diagonal matrices on the space of lattices. See (Cassels and Swinnerton-Dyer 1955) for an implicit description of this correspondence and (Einsiedler and Lindenstrauss 2006; Lindenstrauss 2007; Margulis 1997) for more detail.

The dynamical approach also turned out to be fruitful in studying Diophantine properties of pairs $(\mathbf{f}, v)$ for which the nonplanarity condition fails. Note that obvious examples of non-extremal measures are provided by proper affine subspaces of $\mathbb{R}^n$ whose coefficients are rational or are well enough approximable by rational numbers. On the other hand, it is clear from a Fubini argument that almost all translates of any given subspace are extremal. In (Kleinbock 2003) the method of (Kleinbock and Margulis 1998) was pushed further to produce criteria for the extremality, as well as the strong extremality, of arbitrary affine subspaces $\mathcal{L}$ of $\mathbb{R}^n$. Further, it was shown that if $\mathcal{L}$ is extremal (resp. strongly extremal), then so is any smooth submanifold of $\mathcal{L}$ which is nondegenerate in $\mathcal{L}$ at a.e. point. (The latter property is a straightforward generalization of the definition of non-degeneracy in $\mathbb{R}^n$: a map f is nondegenerate in $\mathcal{L}$ at $\mathbf{x}$ if the linear part of $\mathcal{L}$ is spanned by partial derivatives of $\mathbf{f}$ at $\mathbf{x}$.) In other words, extremality and strong extremality pass from affine subspaces to their nondegenerate submanifolds.

A more precise analysis makes it possible to study Diophantine exponents of measures with supports contained in arbitrary proper affine subspaces of $\mathbb{R}^n$. Namely, in (Kleinbock 2008) it is shown how to compute $\omega(\mathcal{L})$ for any $\mathcal{L}$ and furthermore proved that if μ is a Federer measure on $\mathbb{R}^d$, U an open subset of $\mathbb{R}^d$, and $\mathbf{f} : U \rightarrow \mathbb{R}^n$ a continuous map such that the pair $(\mathbf{f}, v)$ is good and nonplanar in $\mathcal{L}$, then $\omega(\mathbf{f}, v) = \omega(\mathcal{L})$. Here we say, generalizing the definition from section “[Diophantine Approximation with Dependent Quantities: The Set-Up](#),” that $(\mathbf{f}, v)$ is nonplanar in $\mathcal{L}$ if for any ball B with $v(B) > 0$, the $\mathbf{f}$ -image of $B \cap \text{supp } v$ is not contained in any proper affine subspace of $\mathcal{L}$. (It is easy to see that for a smooth map $\mathbf{f} : U \rightarrow \mathcal{L}$, $(\mathbf{f}, \lambda)$ is good and nonplanar in $\mathcal{L}$ whenever f is nondegenerate in $\mathcal{L}$ at a.e. point.) It is worthwhile to point out that these new applications require a strengthening of the measure estimate described at the end of section “[Diophantine](#)

Approximation with Dependent Quantities: The Set-Up”: it was shown in (Kleinbock 2008) that (i) and (ii) would still imply (iii) if ρ in (ii) is replaced by $\rho^{\dim V}$. See (Huang 2022) for the most up-to-date account of what is known about Diophantine exponents of affine subspaces.

Another application concerns badly approximable vectors. Using the dynamical description of the set $BA \subset \mathbb{R}^n$ due to Dani (1985), it turns out to be possible to find badly approximable vectors inside supports of certain measures on $\mathbb{R}^n$. Namely, if a subset K of $\mathbb{R}^n$ supports an absolutely friendly measure, then $BA \cap K$ has Hausdorff dimension not less than the Hausdorff dimension of this measure. In particular, it proves that limit measures of irreducible system of self-similar/self-conformal contractions satisfying the Open Set Condition, such as the middle-third Cantor set on the real line, contain subsets of full Hausdorff dimension consisting of badly approximable vectors. This was established in (Kleinbock and Weiss 2005a) and later independently in (Kristensen et al. 2006) using a different approach. See also Fishman’s work (Fishman 2006) for a stronger result involving winning sets of Schmidt games (the winning property, due to the work of Schmidt (1966), implies full Hausdorff dimension and is stable with respect to countable intersections). Fishman’s result has been significantly generalized in 2012 (Broderick et al. 2012) by showing that BA has the *hyperplane absolute winning* (HAW) property, which is stronger than winning and is inherited by supports of absolutely friendly measures. An even more striking development occurred several years later, when Beresnevich (2015), Yang (2019), and then Beresnevich–Nesharim–Yang (2022) showed that any nondegenerate real analytic curve in $\mathbb{R}^n$ contains a winning set of badly approximable vectors. Moreover, they did it in the context of approximation with weights, thereby giving another proof of Schmidt’s conjecture from (Schmidt 1982), originally proved by Badziahin, Pollington, and Velani (2011). Note that quantitative nondivergence estimates play a crucial role in the argument of (Beresnevich et al. 2022).

It is important to mention that for some special classes of fractals much more can be said: namely,

assuming some invariance properties of the fractal subset, one can aim at divergence Khintchine-type theorems, such as proving that the intersection of BA with the fractal is null with respect to some natural measure supported on the fractal. See (Simmons and Weiss 2019; Khalil and Luethi 2021) for several recent results in this direction. The argument in these papers involves studying nondivergence of random walks on the space of lattices and is based on the work of Benoist and Quint.

The quantitative nondivergence method also has applications to the problem of improving Dirichlet’s Theorem (uniform approximation). For example, in (Kleinbock and Weiss 2008) almost every point of any nondegenerate smooth manifold is proved not to lie in $\mathcal{D}(\psi_{\varepsilon,n/m})$ for small enough ε depending only on the manifold. Some earlier work was done in (Baker 1976; Bugeaud 2002; Davenport and Schmidt 1970) and also in (Kleinbock and Weiss 2005b) for the set of singular vectors. However, a different technique based on Ratner’s Theorem and the Dani–Margulis linearization method (Dani and Margulis 1993) enabled Shah (2009) to prove a much stronger result, namely, the equidistribution of g_t -translates of orbits $\{L_{\mathbf{f}(\mathbf{x})}\mathbb{Z}^{n+1}\}$ for a real analytic nondegenerate $\mathbf{f}$. This was later extended to multiplicative approximation (Shah 2010), smooth nondegenerate maps (Shah and Yang 2022b), affine subspaces (Shah and Yang 2022a; Kleinbock et al. 2021b; Chow and Yang 2019), and submanifolds of $M_{m,n}$ with $\min(m, n) > 1$ (Yang 2016, 2020; Shah and Yang 2020).

It is also worthwhile to mention that a generalization of the measure estimate discussed in section “**Diophantine Approximation with Dependent Quantities: The Set-Up**” was used in (Bernik et al. 2001) to estimate the measure of the set of points x in a ball $B \subset \mathbb{R}^d$ for which the system

$$\begin{cases} |\mathbf{f}(\mathbf{x}) \cdot \mathbf{q} + p| < \varepsilon \\ |\mathbf{f}'(\mathbf{x}) \cdot \mathbf{q}| < \delta \\ |q_i| < Q_i, \quad i = 1, \dots, n, \end{cases}$$

where $\mathbf{f}$ is a smooth nondegenerate map $B \rightarrow \mathbb{R}^n$, has a nonzero integer solution. For that, $L_{\mathbf{f}(\mathbf{x})}$ as in (15) has to be replaced by the matrix

$$\begin{pmatrix} 1 & 0 & \mathbf{f}(\mathbf{x}) \\ 0 & 1 & \mathbf{f}'(\mathbf{x}) \\ 0 & 0 & I_n \end{pmatrix},$$

and therefore (i) and (ii) turn into more complicated conditions, which nevertheless can be checked when $\mathbf{f}$ is smooth and nondegenerate and ν is Lebesgue measure. This has resulted in proving the convergence case of Khintchine–Groshev Theorem for nondegenerate manifolds (Bernik et al. 2001), in both standard and multiplicative versions. The aforementioned estimate was also used in (Beresnevich et al. 2002) for the proof of the divergence case and in (Ghosh 2005, 2006) for establishing the convergence case of Khintchine–Groshev Theorem for affine hyperplanes and their nondegenerate submanifolds. This generalized results obtained by standard methods for the curve (12) by Bernik and Beresnevich (Beresnevich 2002; Bernik 1984).

Note also that in many of the problems mentioned above, the ground field $\mathbb{R}$ can be replaced by $\mathbb{Q}_p$, and in fact several fields can be taken simultaneously, thus giving rise to the S -arithmetic setting where $S = \{p_1, \dots, p_s\}$ is a finite set of normalized valuations of $\mathbb{Q}$, which may or may not include the infinite valuation (cf. [Sprindžuk 1969; Želudevich 1986]). The space of lattices in $\mathbb{R}^{n+1}$ is replaced there by the space of lattices in $\mathbb{Q}_S^{n+1}$, where $\mathbb{Q}_S$ is the product of the fields $\mathbb{R}$ and $\mathbb{Q}_{p_1}, \dots, \mathbb{Q}_{p_s}$. This is the subject of the paper (Kleinbock and Tomanov 2007), where S -arithmetic analogues of many results reviewed in section “[Diophantine Approximation with Dependent Quantities: The Set-Up](#)” have been established. Similarly one can consider versions of the above theorems over local fields of positive characteristic (Ghosh 2007). See also Kleinbock (2004) where Sprindžuk’s solution (Sprindžuk 1969) of the complex case of Mahler’s Conjecture has been generalized (the latter involves studying small values of linear forms with coefficients in $\mathbb{C}$ at real integer points), and (Einsiedler and Kleinbock 2007) which establishes a p -adic analogue of the result of (Einsiedler et al. 2006) on the set of exceptions to Littlewood’s Conjecture.

Finally let us mention another incarnation of Diophantine approximation with dependent quantities, in which the correspondence between approximation and dynamics proved to be helpful: *intrinsic approximation* on manifolds; that is, approximating points on the manifold by rational points lying on the same manifold (as opposed to *ambient* approximation described in the next section). See (Fishman et al. 2018) for some general results and (Kleinbock and Merrill 2015; Fishman et al. 2022; Kleinbock and de Saxcé 2018) for a complete theory when the manifold is a quadric hypersurface; in the latter case, dynamics takes place on the space of lattices in the light cone of the corresponding quadratic form. See also some earlier work by Druţu (2005).

Future Directions

Interactions between ergodic theory and number theory have been rapidly expanding during the last two decades, and the author has no doubts that new applications of dynamics to Diophantine approximation will emerge in the near future. Specializing to the topics discussed in the present paper, it is fair to say that the list of “further results” contained in the previous section is by no means complete, and many even further results are currently in preparation. A lot should be expected from recent developments in Parametric Geometry of Numbers, that is, studying statistics of higher minima of lattices $g_L \mathbb{Z}^k$ (Das et al. 2019; Solan 2021). Let us mention two more directions which have been quite actively pursued in recent years: the problem of *effectivization* of non-effective results in Diophantine approximation, such as Margulis’ solution of the Oppenheim Conjecture (see, e.g., [Lindenstrauss and Margulis 2014; Strömbergsson and Vishe 2020]), and replacing the supremum norm with arbitrary norms in uniform approximation problems (the latter hinges on a careful study of critical loci for those norms, see [Andersen and Duke 2021; Kleinbock and Rao 2022a, b, c]). Several interesting open directions are listed in (Gorodnik 2007), Section 9, in the final sections of papers (Beresnevich et al. 2002; Kleinbock et al. 2004), in the book (Bugeaud 2002), and in surveys by

Frantzikinakis–McCutcheon, Nitica, and Ward in this volume.

Acknowledgments The work on this paper was supported in part by NSF Grants DMS-0239463 and DMS-1900560.

Bibliography

- Baker A (1975) Transcendental number theory. Cambridge University Press, London
- Baker RC (1976) Metric diophantine approximation on manifolds. *J Lond Math Soc* 14:43–48
- Baker RC (1978) Dirichlet's theorem on diophantine approximation. *Math Proc Camb Philos Soc* 83:37–59
- Bekka M, Mayer M (2000) Ergodic theory and topological dynamics of group actions on homogeneous spaces. Cambridge University Press, Cambridge
- Beresnevich V (1999) On approximation of real numbers by real algebraic numbers. *Acta Arith* 90:97–112
- Beresnevich V (2002) A Groshev type theorem for convergence on manifolds. *Acta Math Hungar* 94:99–130
- Beresnevich V, Velani S (2007) A note on simultaneous Diophantine approximation on planar curves. *Ann Math* 337:769–796
- Beresnevich V, Bernik VI, Kleinbock D, Margulis GA (2002) Metric Diophantine approximation: the Khintchine-Groshev theorem for nondegenerate manifolds. *Moscow Math J* 2:203–225
- Beresnevich V, Dickinson H, Velani S (2006) Measure theoretic laws for lim sup sets. *Memoirs Am Math Soc* 179:1–91
- Beresnevich V, Dickinson H, Velani S (2007) Diophantine approximation on planar curves and the distribution of rational points. *Ann Math* 166:367–426
- Bernik VI (1984) A proof of Baker's conjecture in the metric theory of transcendental numbers. *Dokl Akad Nauk SSSR* 277:1036–1039
- Bernik VI, Dodson MM (1999) Metric Diophantine approximation on manifolds. Cambridge University Press, Cambridge
- Bernik VI, Kleinbock D, Margulis GA (2001) Khintchine-type theorems on manifolds: convergence case for standard and multiplicative versions. *Int Math Res Notices* 2001:453–486
- Besicovitch AS (1929) On linear sets of points of fractional dimensions. *Ann Math* 101:161–193
- Bugeaud Y (2002) Approximation by algebraic integers and Hausdorff dimension. *J Lond Math Soc* 65:547–559
- Bugeaud Y (2004) Approximation by algebraic numbers. Cambridge University Press, Cambridge
- Cassels JWS (1957) An introduction to Diophantine approximation. Cambridge University Press, New York
- Cassels JWS, Swinnerton-Dyer H (1955) On the product of three homogeneous linear forms and the indefinite ternary quadratic forms. *Philos Trans R Soc Lond* 248:73–96
- Dani SG (1979) On invariant measures, minimal sets and a lemma of Margulis. *Invent Math* 51:239–260
- Dani SG (1985) Divergent trajectories of flows on homogeneous spaces and diophantine approximation. *J Reine Angew Math* 359:55–89
- Dani SG (1986) On orbits of unipotent flows on homogeneous spaces. II. *Ergodic Theory Dyn Syst* 6:167–182
- Dani SG, Margulis GA (1993) Limit distributions of orbits of unipotent flows and values of quadratic forms. In: IM Gelfand Seminar. American Mathematical Society, Providence, pp 91–137
- Davenport H, Schmidt WM (1969/1970) Dirichlet's theorem on diophantine approximation. II. *Acta Arith* 16:413–424
- Davenport H, Schmidt WM (1970) Dirichlet's theorem on diophantine approximation. In: *Symposia Mathematica*. INDAM, Rome, pp 113–132
- Ding J (1994) A proof of a conjecture of C. L. Siegel. *J Number Theory* 46:1–11
- Dodson MM (1992) Hausdorff dimension, lower order and Khintchine's theorem in metric Diophantine approximation. *J Reine Angew Math* 432:69–76
- Dodson MM (1993) Geometric and probabilistic ideas in the metric theory of Diophantine approximations. *Uspekhi Mat Nauk* 48:77–106
- Druţu C (2005) Diophantine approximation on rational quadrics. *Ann Math* 333:405–469
- Duffin RJ, Schaeffer AC (1941) Khintchine's problem in metric Diophantine approximation. *Duke Math J* 8:243–255
- Einsiedler M, Kleinbock D (2007) Measure rigidity and p -adic Littlewood-type problems. *Compos Math* 143:689–702
- Einsiedler M, Lindenstrauss E (2006) Diagonalizable flows on locally homogeneous spaces and number theory. In: *Proceedings of the international congress of mathematicians*. Eur Math Soc, Zurich, pp 1731–1759
- Einsiedler M, Katok A, Lindenstrauss E (2006) Invariant measures and the set of exceptions to Littlewood's conjecture. *Ann Math* 164:513–560
- Eskin A (1998) Counting problems and semisimple groups. In: *Proceedings of the international congress of mathematicians*. Doc Math, Berlin, pp 539–552
- Eskin A, Margulis GA, Mozes S (1998) Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture. *Ann Math* 147:93–141
- Eskin A, Margulis GA, Mozes S (2005) Quadratic forms of signature (2,2) and eigenvalue spacings on rectangular 2-tori. *Ann Math* 161:679–725
- Fishman L (2006) Schmidt's games on certain fractals. *Israel S Math* (to appear)
- Gallagher P (1962) Metric simultaneous diophantine approximation. *J Lond Math Soc* 37:387–390
- Ghosh A (2005) A Khintchine type theorem for hyperplanes. *J Lond Math Soc* 72:293–304
- Ghosh A (2006) Dynamics on homogeneous spaces and Diophantine approximation on manifolds. PhD thesis, Brandeis University, Waltham
- Ghosh A (2007) Metric Diophantine approximation over a local field of positive characteristic. *J Number Theory* 124:454–469
- Gorodnik A (2007) Open problems in dynamics and related fields. *J Mod Dyn* 1:1–35

- Groshev AV (1938) Une théorème sur les systèmes des formes linéaires. Dokl Akad Nauk SSSR 9:151–152
- Harman G (1998) Metric number theory. Clarendon Press/Oxford University Press, New York
- Hutchinson JE (1981) Fractals and self-similarity. Indiana Univ Math J 30:713–747
- Jarník V (1928–9) Zur metrischen Theorie der diophantischen Approximationen. Prace Mat-Fiz 36: 91–106
- Jarník V (1929) Diophantischen Approximationen und Hausdorffsches Mass. Mat Sb 36:371–382
- Khanin K, Lopes-Dias L, Marklof J (2007) Multidimensional continued fractions, dynamical renormalization and KAM theory. Commun Math Phys 270:197–231
- Khintchine A (1924) Einige Satze uber Kettenbrüche, mit Anwendungen auf die Theorie der Diophantischen Approximationen. Math Ann 92:115–125
- Khintchine A (1963) Continued fractions. P Noordhoff Ltd, Groningen
- Kleinbock D (2001) Some applications of homogeneous dynamics to number theory. In: Smooth ergodic theory and its applications. American Mathematical Society, Providence, pp 639–660
- Kleinbock D (2003) Extremal subspaces and their submanifolds. Geom Funct Anal 13:437–466
- Kleinbock D (2004) Baker-Sprindžuk conjectures for complex analytic manifolds. In: Algebraic groups and arithmetic. Tata Inst Fund Res, Mumbai, pp 539–553
- Kleinbock D (2008) An extension of quantitative non-divergence and applications to Diophantine exponents. Trans AMS 360:6497–6523
- Kleinbock D, Margulis GA (1996) Bounded orbits of non-quasiunipotent flows on homogeneous spaces. In: Sinaï's Moscow seminar on dynamical systems. American Mathematical Society, Providence, pp 141–172
- Kleinbock D, Margulis GA (1998) Flows on homogeneous spaces and Diophantine approximation on manifolds. Ann Math 148:339–360
- Kleinbock D, Margulis GA (1999) Logarithm laws for flows on homogeneous spaces. Invent Math 138: 451–494
- Kleinbock D, Tomanov G (2007) Flows on S^1 ? Arithmetic homogeneous spaces and applications to metric Diophantine approximation. Comment Math Helv 82: 519–581
- Kleinbock D, Weiss B (2005a) Badly approximable vectors on fractals. Israel J Math 149:137–170
- Kleinbock D, Weiss B (2005b) Friendly measures, homogeneous flows and singular vectors. In: Algebraic and topological dynamics. American Mathematical Society, Providence, pp 281–292
- Kleinbock D, Weiss B (2008) Dirichlet's theorem on diophantine approximation and homogeneous flows. J Mod Dyn 2:43–62
- Kleinbock D, Shah N, Starkov A (2002) Dynamics of subgroup actions on homogeneous spaces of Lie groups and applications to number theory. In: Handbook on dynamical systems, vol 1A. Elsevier Science, North Holland, Amsterdam, pp 813–930
- Kleinbock D, Lindenstrauss E, Weiss B (2004) On fractal measures and diophantine approximation. Selecta Math 10:479–523
- Kontsevich M, Suhov Y (1999) Statistics of Klein polyhedra and multidimensional continued fractions. In: Pseudoperiodic topology. American Mathematical Society, Providence, pp 9–27
- Kristensen S, Thorn R, Velani S (2006) Diophantine approximation and badly approximable sets. Adv Math 203:132–169
- Lagarias JC (1994) Geodesic multidimensional continued fractions. Proc Lond Math Soc 69:464–488
- Lindenstrauss E (2007) Some examples how to use measure classification in number theory. In: Equidistribution in number theory, an introduction. Springer, Dordrecht, pp 261–303
- Mahler K (1932) Über das Mass der Menge aller S -Zahlen. Math Ann 106:131–139
- Margulis GA (1975) On the action of unipotent groups in the space of lattices. In: Lie groups and their representations (Budapest, 1971). Halsted, New York, pp 365–370
- Margulis GA (1989) Discrete subgroups and ergodic theory. In: Number theory, trace formulas and discrete groups (Oslo, 1987). Academic, Boston, pp 377–398
- Margulis GA (1997) Oppenheim conjecture. In: Fields medalists' lectures. World Sci Publishing, River Edge, pp 272–327
- Margulis GA (2002) Diophantine approximation, lattices and flows on homogeneous spaces. In: A panorama of number theory or the view from Baker's garden. Cambridge University Press, Cambridge, pp 280–310
- Mauldin D, Urbański M (1996) Dimensions and measures in infinite iterated function systems. Proc Lond Math Soc 73:105–154
- Mohammadi A, Salehi Golsefidy A (2008) S -arithmetic Khintchine-type theorem. Preprint
- Moore CC (1966) Ergodicity of flows on homogeneous spaces. Am J Math 88:154–178
- Roth KF (1955) Rational approximations to algebraic numbers. Mathematika 2:1–20
- Schmidt WM (1960) A metrical theorem in diophantine approximation. Can J Math 12:619–631
- Schmidt WM (1964) Metrische Satze uber simultane Approximation abhangiger Grosen. Monatsh Math 63:154–166
- Schmidt WM (1969) Badly approximable systems of linear forms. J Number Theory 1:139–154
- Schmidt WM (1972) Norm form equations. Ann Math 96: 526–551
- Schmidt WM (1980) Diophantine approximation. Springer, Berlin
- Sheingorn M (1993) Continued fractions and congruence subgroup geodesics. In: Number theory with an emphasis on the Markoff spectrum (Provo, UT, 1991). Dekker, New York, pp 239–254
- Sprindžuk VG (1964) More on Mahler's conjecture. Dokl Akad Nauk SSSR 155:54–56
- Sprindžuk VG (1969) Mahler's problem in metric number theory. American Mathematical Society, Providence

- Sprindžuk VG (1979) Metric theory of Diophantine approximations. VH Winston & Sons, Washington, DC
- Sprindžuk VG (1980) Achievements and problems of the theory of Diophantine approximations. *Uspekhi Mat Nauk* 35:3–68
- Starkov A (2000) Dynamical systems on homogeneous spaces. American Mathematical Society, Providence
- Stratmann B, Urbanski M (2006) Diophantine extremality of the Patterson measure. *Math Proc Camb Philos Soc* 140:297–304
- Sullivan D (1982) Disjoint spheres, approximation by imaginary quadratic numbers, and the logarithm law for geodesics. *Acta Math* 149:215–237
- Urbanski M (2005) Diophantine approximation of self-conformal measures. *J Number Theory* 110:219–235
- Želudevič F (1986) Simultane diophantische Approximationen abhängiger Grossen in mehreren Metriken. *Acta Arith* 46:285–296
- ### New List
- Aka M, Breuillard E, Rosenzweig L, de Saxcé N (2018) Diophantine approximation of matrices and Lie groups. *Geom Funct Anal* 28:1–57
- Andersen N, Duke W (2021) On a theorem of Davenport and Schmidt. *Acta Arith* 198:37–75
- Athreya J, Ghosh A, Tseng J (2015) Spiraling of approximations and spherical averages of Siegel transforms. *J Lond Math Soc* 91:383–404
- Badziahin D, Pollington A, Velani S (2011) On a problem in simultaneous Diophantine approximation: Schmidt's conjecture. *Ann Math* 174:1837–1883
- Beresnevich V (2015) Badly approximable points on manifolds. *Inv Math* 202:1199–1240
- Beresnevich V, Kleinbock D (2022) Quantitative non-divergence and Diophantine approximation on manifolds. In: Dynamics, geometry, number theory: the impact of Margulis on modern mathematics. University of Chicago Press, Chicago, pp 303–341
- Beresnevich V, Kleinbock D, Margulis GA (2015) Non-planarity and metric Diophantine approximation for systems of linear forms. *J Number Theory Bordeaux* 27:1–31
- Beresnevich V, Guan L, Marnat A, Ramirez F, Velani S (2020) Dirichlet is not just bad and singular. Preprint arXiv:2008.04043
- Beresnevich V, Nesharim E, Yang L (2022) Winning property of badly approximable points on curves. *Duke Math J*, to appear
- Björklund M, Gorodnik A (2019) Central limit theorems for Diophantine approximants. *Math Ann* 374:1371–1437
- Broderick R, Kleinbock D, Fishman L, Reich A, Weiss B (2012) The set of badly approximable vectors is strongly C^1 incompressible. *Math Proc Camb Philos Soc* 153:319–339
- Cheung Y (2011) Hausdorff dimension of the set of singular pairs. *Ann Math* 173:127–167
- Cheung Y, Chevallier N (2016) Hausdorff dimension of singular vectors. *Duke Math J* 165:2273–2329
- Chow S, Yang L (2019) An effective Ratner equidistribution theorem for multiplicative Diophantine approximation on planar lines. Preprint arXiv:1902.06081
- Das T, Fishman L, Simmons D, Urbański M (2018) Extremality and dynamically defined measures, part I: Diophantine properties of quasi-decaying measures. *Selecta Math* 24:2165–2206
- Das T, Fishman L, Simmons D, Urbański M (2019) A variational principle in the parametric geometry of numbers. Preprint arXiv:1901.06602
- Dolgopyat D, Fayad B, Vinogradov I (2017) Central limit theorems for simultaneous Diophantine approximations. *J Éc polytech Math* 4:1–36
- Eskin A, Mozes S (2022) Margulis functions and their applications. In: Dynamics, geometry, number theory: the impact of Margulis on modern mathematics. University of Chicago Press, Chicago, pp 342–361
- Fishman L, Kleinbock D, Merrill K, Simmons D (2018) Intrinsic Diophantine approximation on manifolds: general theory. *TAMS* 370:577–599
- Fishman L, Kleinbock D, Merrill K, Simmons D (2022) Intrinsic Diophantine approximation on quadric hypersurfaces. *JEMS* 24:1045–1101
- Huang JJ (2022) Extremal affine subspaces and Khintchine-Jarník type theorems. Preprint, arXiv:2208.04255
- Kadyrov S, Kleinbock D, Lindenstrauss E, Margulis GA (2017) Singular systems of linear forms and non-escape of mass in the space of lattices. *J Anal Math* 133:253–277
- Khalil O, Luethi M (2021) Random walks, spectral gaps, and Khintchine's theorem on fractals. Preprint arXiv:2101.05797
- Kim T, Kim W (2022) Hausdorff measure of sets of Dirichlet non-improvable affine forms. *Adv Math* 403:Paper No. 108353
- Kleinbock D (2010) Quantitative nondivergence and its Diophantine applications. Lecture notes for Clay Institute Summer School (Pisa 2007). In: Homogeneous flows, moduli spaces and arithmetic, *Clay Math Proc*, pp 131–153
- Kleinbock D, de Saxcé N (2018) Rational approximation on quadrics: a simplex lemma and its consequences. *Enseign Math* 64:459–476
- Kleinbock D, Merrill K (2015) Rational approximation on spheres. *Israel J Math* 209:293–322
- Kleinbock D, Mirzadeh S (2020) On the dimension drop conjecture for diagonal flows on the space of lattices. *Adv Math*, to appear
- Kleinbock D, Rao A (2022a) A zero-one law for uniform Diophantine approximation in Euclidean norm. *Int Math Res Not* 2022:5617–5657
- Kleinbock D, Rao A (2022b) Abundance of Dirichlet-improvable pairs with respect to arbitrary norms. *Mosc J Comb Number Theory* 11:97–114
- Kleinbock D, Rao A (2022c) Weighted uniform Diophantine approximation of systems of linear forms. *Pure Appl Math Q* 18:1095–1112

- Kleinbock D, Wadleigh N (2018) A zero-one law for improvements to Dirichlet's theorem. *PAMS* 146: 1833–1844
- Kleinbock D, Wadleigh N (2019) An inhomogeneous Dirichlet theorem via shrinking targets. *Compos Math* 155:1402–1423
- Kleinbock D, Weiss B (2013) Modified Schmidt games and a conjecture of Margulis. *J Mod Dyn* 7:429–460
- Kleinbock D, Margulis GA, Wang J (2010) Metric Diophantine approximation for systems of linear forms via dynamics. *Int J Number Theory* 6:1139–1168
- Kleinbock D, Strombergsson A, Yu S (2021a) Measure estimates and improvements to Dirichlet's theorem. *Proc. London Math Soc*, to appear
- Kleinbock D, de Saxcé N, Shah N, Yang P (2021b) Equidistribution in the space of 3-lattices and Dirichlet-improvable vectors on planar lines. *Ann Sc Norm Super Pisa*, to appear
- Lindenstrauss E, Margulis G (2014) Effective estimates on indefinite ternary forms. *Israel J Math* 203:445–499
- Roy D (2015) On Schmidt and Summerer parametric geometry of numbers. *Ann Math* 182:739–786
- Schmidt WM (1966) On badly approximable numbers and certain games. *TAMS* 123:78–199
- Schmidt WM (1982) Open problems in Diophantine approximation. In: *Diophantine approximations and transcendental numbers*. *Progr. Math.*, 31, Birkhäuser Boston, Boston, MA, pp 271–287
- Schmidt WM, Summerer L (2013) Diophantine approximation and parametric geometry of numbers. *Monatsh Math* 169:51–104
- Shah N (2009) Equidistribution of expanding translates of curves and Dirichlet's theorem on Diophantine approximation. *Invent Math* 177:509–532
- Shah N (2010) Expanding translates of curves and Dirichlet-Minkowski theorem on linear forms. *JAMS* 23:563–589
- Shah N, Yang L (2020) Equidistribution of curves in homogeneous spaces and Dirichlet's approximation theorem for matrices. *Discrete Contin Dyn Syst* 40: 5247–5287
- Shah N, Yang P (2022a) Equidistribution of expanding degenerate manifolds in the space of lattices. Preprint arXiv:2112.13952
- Shah N, Yang P (2022b) Equidistribution of non-uniformly stretching translates of shrinking smooth curves and weighted Dirichlet approximation. Preprint arXiv:2204.03194
- Shapira U, Weiss B (2022) Geometric and arithmetic aspects of approximation vectors. Preprint arXiv:2206.05329
- Simmons D, Weiss B (2019) Random walks on homogeneous spaces and Diophantine approximation on fractals. *Invent Math* 216:337–394
- Solan O (2021) Parametric geometry of numbers with general flow. Preprint, arXiv:2106.01707
- Strömbergsson A, Vishe P (2020) An effective equidistribution result for $SL(2, \mathbb{R}) \ltimes (\mathbb{R}^2)^{\oplus k}$ and application to inhomogeneous quadratic forms. *J Lond Math Soc* 102:143–204
- Yang L (2016) Equidistribution of expanding curves in homogeneous spaces and Diophantine approximation on square matrices. *PAMS* 144:5291–5308
- Yang L (2019) Badly approximable points on manifolds and unipotent orbits in homogeneous spaces. *Geom Funct Anal* 29:1194–1234
- Yang P (2020) Equidistribution of expanding translates of curves and Diophantine approximation on matrices. *Invent Math* 220:909–948



Ergodic Theory: Rigidity

Viorel Nițică

West Chester University, West Chester, PA, USA

Institute of Mathematics, Bucharest, Romania

Article Outline

Glossary

Definition of the Subject

Introduction

Basic Definitions and Examples

Differentiable Rigidity

Local Rigidity

Global Rigidity

Measure Rigidity

Future Directions

Bibliography

Glossary

C^k Conjugacy Two diffeomorphisms ϕ_1, ϕ_2 acting on the same manifold M are said to be C^k -conjugated if there exists a C^k diffeomorphism h of M such that $\phi_1 = h^{-1} \circ \phi_2 \circ h$. The diffeomorphism h is called C^k conjugacy.

Conjugacy Two elements g_1, g_2 in a group G are said to be conjugated if there exists an element $h \in G$ such that $g_1 = h^{-1}g_2h$. The element h is called conjugacy.

Differentiable rigidity Differentiable rigidity refers to finding invariants to the differentiable conjugacy of dynamical systems, and, more general, group actions.

Global rigidity Global rigidity refers to the classification of all group actions on manifolds satisfying certain conditions.

Lattice A lattice in a Lie group is a discrete subgroup of finite covolume.

Local rigidity Local rigidity refers to the study of perturbations of homomorphisms from discrete or continuous groups into diffeomorphism groups.

Measure rigidity Measure rigidity refers to the study of invariant measures for actions of abelian groups and semigroups.

Definition of the Subject

As one can see from this volume, chaotic behavior of complex dynamical systems is prevalent in nature and in large classes of transformations. Rigidity theory can be viewed as the counterpart to the generic theory of dynamical systems which often investigates chaotic dynamics for a typical transformation belonging to a large class. In rigidity one is interested in finding obstructions to chaotic, or generic, behavior. This often leads to rather unexpected classification results. As such, rigidity in dynamics and ergodic theory is difficult to define precisely and the best approach to this subject is to study various results and themes that developed so far. A classification is offered below in local, global, differentiable and measurable rigidity. One should note that all branches are strongly intertwined and, at this stage of the development of the subject, it is difficult to separate them.

Rigidity is a well developed and prominent topic in modern mathematics. Historically, rigidity has two main origins, one coming from the study of lattices in semi-simple Lie groups, and one coming from the theory of hyperbolic dynamical systems. From the start, ergodic theory was an important tool used to prove rigidity results, and a strong interdependence developed between these fields. Many times a result in rigidity is obtained by combining techniques from the theory of lattices in Lie groups with techniques from hyperbolic dynamical systems and ergodic theory. Among other mathematical disciplines using results and contributing to this field one can mention representation theory, smooth, continuous and measurable dynamics, harmonic and spectral analysis, partial differential equations, differential geometry, and number theory. Additional details about the appearance of rigidity in ergodic theory

as well as definitions for some terminology used in the sequel can be found in the articles ► [“Ergodic Theory on Homogeneous Spaces and Metric Number Theory”](#) by Kleinbock, ► [“Ergodic Theory: Recurrence”](#) by Frantzikinakis, McCutcheon, and ► [“Ergodic Theory: Interactions with Combinatorics and Number Theory”](#) by Ward. The theory of hyperbolic dynamics is presented in the entry “Hyperbolic Dynamical Systems” by Viana and in the entry ► [“Smooth Ergodic Theory”](#) by Wilkinson.

Introduction

The first results about classification of lattices in semi-simple Lie groups were local, aimed at trying to understand the space of small perturbations of a given linear representation. A major contributor was Weil (1960, 1962, 1964), who proved local rigidity of linear representations for large classes of groups, in particular lattices. Another breakthrough was the contribution of Kazhdan (1967), who introduced property (T), allowing to show that large classes of lattices are finitely generated. Rigidity theory matured due to the remarkable global rigidity results obtained by Mostow (1973) and Margulis (1991), leading to a complete classification of lattices in large classes of semi-simple Lie groups.

Briefly, a hyperbolic (Anosov) dynamical system is one that exhibits strong expansion and contraction along complementary directions. An early contribution introducing this class of objects is the paper of Smale (1967), in which basic examples and techniques are introduced. A breakthrough came with the results of Anosov (1967), who proved structural stability of the Anosov systems and ergodicity of the geodesic flow on a manifold of negative curvature. Motivated by questions arising in mathematical physics, chaos theory and other areas, hyperbolic dynamics emerged as one of the major fields of contemporary mathematics. From the beginning, a major unsolved problem in the field was the classification of Anosov diffeomorphisms and flows.

In the 80s a change in philosophy occurred, partially motivated by a program introduced by

Zimmer (1987). The goal of the program was to classify the smooth actions of higher rank semi-simple Lie groups and of their (irreducible) lattices on compact manifolds. It was expected that any such lattice action that preserves a smooth volume form and is ergodic can be reduced to one of the following standard models: isometric actions, linear actions on infranil manifolds, and left translations on compact homogeneous spaces. This original conjecture was disproved by Katok, Lewis (see (Katok and Lewis 1996)): by blowing up a linear nilmanifold-action at some fixed points they exhibit real-analytic, volume preserving, ergodic lattice actions on manifolds with complicated topology.

Nevertheless, imposing extra assumptions on a higher rank action, for example the existence of some hyperbolicity, allows local and global classification results. The concept of *Anosov action*, that is, an action that contains at least one Anosov diffeomorphism, was introduced for general groups by Pugh, Shub (1972). The significant differences between the classical $\mathbb{Z}$ and $\mathbb{R}$ cases and those of higher rank lattices, or at a more basic level, of higher rank abelian groups, went unnoticed for a while. The surge of activity in the 80s allowed these differences to surface: for lattices in the work of Hurder, Katok, Lewis, Zimmer (see (Hurder 1992; Katok and Lewis 1991, 1996; Katok et al. 1996)); and for higher rank abelian groups in the work of Katok, Lewis (1991) and Katok, Spatzier (1994a). As observed in these papers, local and global rigidity are typical for such Anosov actions. This generated additional research which is summarized in section [“Local Rigidity”](#) and [“Global Rigidity”](#).

Differentiable rigidity is covered in section [“Differentiable Rigidity”](#). An interesting problem is to find moduli for the C^k conjugacy, $k \geq 1$, of Anosov diffeomorphisms and flows. This was tackled so far only for low dimensional cases ($n = 2$ for diffeomorphisms and $n = 3$ for flows). Another direction that can be included here refers to finding obstructions for higher transverse regularity of the stable/unstable foliation of a hyperbolic system. A spin-off of the research done so far, which is of high interest by itself, and has applications to local and global rigidity,

consists of results lifting the regularity of solutions of cohomological equations over hyperbolic systems. In turn, these results motivated a more careful study of analytic questions about lifting the regularity of real valued continuous functions that enjoy higher regularity along webs of foliations. We also include in this section rigidity results for cocycles over higher rank abelian actions. These are crucial to the proof of local rigidity of higher rank abelian group actions. A more detailed presentation of the material relevant to differentiable rigidity can be found in the forthcoming monograph (Katok and Nițică).

Measure rigidity refers to the study of invariant measures under actions of abelian groups and semi-groups. If the actions are hyperbolic, higher-rank, and satisfy natural algebraic and irreducibility assumptions, one expects the invariant measures to be rare. This direction was started by a question of Furstenberg, asking if any nonatomic probability measure on the circle, invariant and ergodic under multiplications by 2 and 3, is the Lebesgue measure. An early contribution is that of Rudolph (1990), who answered positively if the action has an element of strictly positive entropy. Katok, Spatzier (1996) extended the question to more general higher rank abelian actions, such as actions by linear automorphisms of tori and Weyl chamber flows. A related direction is the study of the invariant sets and measures under the action of horocycle flows, where important progress was made by Ratner (1991a, b) and earlier by Margulis (1989, 1997; Dani and Margulis 1990). An application of these results present in the last papers is the proof of the long standing Oppenheim's conjecture, about the density of the values of the quadratic forms at integer points. Recent developments due to and Einsiedler, Katok, Lindenstrauss (2006) give a partial answer to another outstanding conjecture in number theory, Littlewood's conjecture, and emphasize measure rigidity as one of a more promising directions in rigidity. More details are shown in section "Measure Rigidity".

Four other recent surveys of rigidity theory, each one with a fair amount of overlap but also complementary in part to the present one, that discuss various aspects of the field and its significance are written by Fisher (2006), Lindenstrauss

(2005), Nițică, Török (2002), and Spatzier (1995). Among these, (Fisher 2006) concentrates mostly on local and global rigidity, (Nițică and Török 2002) on differentiable rigidity, (Lindenstrauss 2005) on measure rigidity, and (Spatzier 1995) gives a general overview.

Here is a word of caution for the reader. Many times, instead of the most general results, we present an example that contains the essence of what is available. Also, several important facts that should have been included, are left out. This is because stating complete results would require more space than allocated to this material. The limited knowledge of the author also plays a role here. He apologizes for any obvious omissions and hopes that the bibliography will help fill the gaps.

Basic Definitions and Examples

A detailed introduction to the theory of Anosov systems and hyperbolic dynamics is given in the monograph (Katok and Hasselblatt 1995). The proofs of the basic results for diffeomorphisms stated below can be found there. The proofs for flows are similar. Surveys about hyperbolic dynamics in this volume are the article "Hyperbolic Dynamical Systems" by Viana and the entry ▶ "Smooth Ergodic Theory" by Wilkinson.

Consider a compact differentiable manifold M and $f: M \rightarrow M$ a C^1 diffeomorphism. Let TM be the tangent bundle of M , and $Df: TM \rightarrow TM$ be the derivative of f . The map f is said to be an *Anosov diffeomorphism* if there is a smooth Riemannian metric $\|\cdot\|$ on M , which induces a metric d_M called *adapted*, a number $\lambda \in (0, 1)$, and a continuous Df invariant splitting $TM = E^s \oplus E^u$ such that

$$\|Df v\| \leq \lambda \|v\|, v \in E^s, \quad \|Df^{-1} v\| \leq \lambda \|v\|, v \in E^u,$$

For each $x \in M$ there is a pair of embedded C^1 discs $W_{\text{loc}}^s(x), W_{\text{loc}}^u(x)$, called the local stable manifold and the local unstable manifold at x , respectively, such that:

1. $T_x W_{\text{loc}}^s(x) = E^s(x)$, $T_x W_{\text{loc}}^u(x) = E^u(x)$;
2. $f(W_{\text{loc}}^s(x)) \subset W_{\text{loc}}^s(fx)$, $f^{-1}(W_{\text{loc}}^u(x)) \subset W_{\text{loc}}^u(f^{-1}x)$;

3. For any $\mu \in (\lambda, 1)$, there exists a constant $C > 0$ such that for all $n \in \mathbb{N}$,

$$d_M(f^n x, f^n y) \leq C\mu^n d_M(x, y), \text{ for } y \in W_{\text{loc}}^s(x),$$

$$d_M(f^{-n} x, f^{-n} y) \leq C\mu^n d_M(x, y), \text{ for } y \in W_{\text{loc}}^u(x).$$

The local stable (unstable) manifolds can be extended to global stable (unstable) manifolds $W^s(x)$ and $W^u(x)$ which are well defined and smoothly injectively immersed. These global manifolds are the leaves of global foliations W^s and W^u of M . In general, these foliations are only continuous, but their leaves are differentiable.

Let $\phi: \mathbb{R} \times M \rightarrow M$ be a C^1 flow. The flow ϕ is said to be an *Anosov flow* if there is a Riemannian metric $\|\cdot\|$ on M , a constant $0 < \lambda < 1$, and a continuous Df invariant splitting $TM = E^s \oplus E^0 \oplus E^u$ such that for all $x \in M$ and $t > 0$:

1. $\frac{d}{dt}\big|_{t=0} \phi^t \in E_x^c \setminus \{0\}$, $\dim E_x^c = 1$,
2. $\|D\phi^t v\| \leq \lambda^t \|v\|$, $v \in E^s$,
3. $\|D\phi^{-t} v\| \leq \lambda^t \|v\|$, $v \in E^u$.

For each $x \in M$ there is a pair of embedded C^1 discs $W_{\text{loc}}^s(x), W_{\text{loc}}^u(x)$, called the local (strong) stable manifold and the local (strong) unstable manifold at x , respectively, such that:

1. $T_x W_{\text{loc}}^s(x) = E^s(x)$, $T_x W_{\text{loc}}^u(x) = E^u(x)$;
2. $\phi^t(W_{\text{loc}}^s(x)) \subset W_{\text{loc}}^s(\phi^t x)$,
 $\phi^{-t}(W_{\text{loc}}^u(x)) \subset W_{\text{loc}}^u(\phi^{-t} x)$ for $t > 0$;
3. For any $\mu \in (\lambda, 1)$, there exists a constant $C > 0$ such that for all $n \in \mathbb{N}$,

$$d_M(\phi^t x, \phi^t y) \leq C\mu^t d_M(x, y),$$

for $y \in W_{\text{loc}}^s(x), t > 0$,

$$d_M(\phi^{-t} x, \phi^{-t} y) \leq C\mu^t d_M(x, y),$$

for $y \in W_{\text{loc}}^u(x), t > 0$.

The local stable (unstable) manifolds can be extended to global stable (unstable) manifolds $W^s(x)$ and $W^u(x)$. These global manifolds are the leaves of global foliations W^s and W^u of M . One can also define weak stable and weak unstable foliations with leaves given by $W^{cs}(x) = \bigcup_{t \in \mathbb{R}} (W^s(x))$ and $W^{cu}(x) = \bigcup_{t \in \mathbb{R}} (W^u(x))$, which have as tangent

distributions $E^{cs} = E^c \oplus E^s$ and $E^{cu} = E^c \oplus E^u$. In general, all these foliations are only continuous, but their leaves are differentiable.

Any Anosov diffeomorphism is *structurally stable*, that is, any C^1 diffeomorphism that is C^1 close to an Anosov diffeomorphism is topologically conjugate to the unperturbed one via a Hölder homeomorphism. An Anosov flow is structurally stable in the orbit equivalence sense: any C^1 small perturbation of an Anosov flow has the orbit foliation topologically conjugate via a Hölder homeomorphism to the orbit foliation of the unperturbed flow.

Let $SL(n, \mathbb{R})$ be the group of all n -dimensional square matrices with real valued entries of determinant 1. Let $SL(n, \mathbb{Z}) \subset SL(n, \mathbb{R})$ be the subgroup with integer entries. Basic examples of Anosov diffeomorphisms are automorphisms of the n -torus $\mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n$ induced by hyperbolic matrices in $SL(n, \mathbb{Z})$. A hyperbolic matrix is one that has only nonzero eigenvalues, all away in absolute value from 1. A specific example of such matrix in $SL(2, \mathbb{Z})$ is

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Basic examples of Anosov flows are given by the geodesic flows of surfaces of constant negative curvature. The unitary bundle of such a surface can be realized as $M = \Gamma \backslash PSL(2, \mathbb{R})$, where $PSL(2, \mathbb{R}) = SL(2, \mathbb{R}) / \{\pm 1\}$ and Γ is a cocompact lattice in $PSL(2, \mathbb{R})$. The action of the geodesic flow on M is induced by right multiplication with elements in the diagonal one-parameter subgroup

$$\left\{ \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}, t \in \mathbb{R} \right\}.$$

A related transformation, which is not hyperbolic, but will be of interest in this presentation, is the *horocycle flow* induced by right multiplication on M by elements in the one parameter subgroup

$$\left\{ \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, t \in \mathbb{R} \right\}.$$

Of interest in this survey are also actions of more general groups than $\mathbb{Z}$ and $\mathbb{R}$. Typical examples of

higher rank $\mathbb{Z}^k$ Anosov actions are constructed on tori using groups of units in number fields. See (Katok et al. 2002) for more details about this construction. A particular example of Anosov $\mathbb{Z}^2$ -action on $\mathbb{T}^3$ is induced by the hyperbolic matrices:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 8 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 1 & 8 & 4 \end{pmatrix}.$$

One can check, by looking at the eigenvalues, that A and B are not multiples of the same matrix. Moreover, A and B commute.

Typical examples of higher rank Anosov $\mathbb{R}^k$ -actions are given by Weyl chamber flows, which we now describe using some notions from the theory of Lie groups. A good reference for the background in Lie group theory necessary here is the book of Helgason (1978). Note that for a hyperbolic element of such an action the center distribution is k -dimensional and coincides with the tangent distribution to the orbit foliation of $\mathbb{R}^k$.

Let G be a semi-simple connected real Lie group of the noncompact type, with Lie algebra $\mathfrak{g}$. Let $K \subset G$ be a maximal compact subgroup that gives a Cartan decomposition $\mathfrak{g} = \mathfrak{k} + \mathfrak{p}$, where $\mathfrak{k}$ is the Lie algebra of K and $\mathfrak{p}$ is the orthogonal complement of $\mathfrak{k}$ with respect to the Killing form of $\mathfrak{g}$. Let $\mathfrak{a} \subset \mathfrak{p}$ be a maximal abelian subalgebra and $A = \exp \mathfrak{a}$ be the corresponding subgroup. The simultaneous diagonalization of $ad_{\mathfrak{g}}(\mathfrak{a})$ gives the decomposition

$$\mathfrak{g} = \mathfrak{g} + \sum_{\lambda \in \Lambda} \mathfrak{g}_{\lambda}, \quad \mathfrak{g}_0 = \mathfrak{a} + \mathfrak{m},$$

where Λ is the set of restricted roots. The spaces $\mathfrak{g}_{\lambda}$ are called *root spaces*. A point $H \in \mathfrak{a}$ is called *regular* if $\lambda(H) \neq 0$ for all $\lambda \in \Lambda$. Otherwise it is called *singular*. The set of regular elements consists of the complement of a union of finitely many hyperplanes. Its components are cones in a called *Weyl chambers*. The faces of the Weyl chambers are called *Weyl chamber walls*.

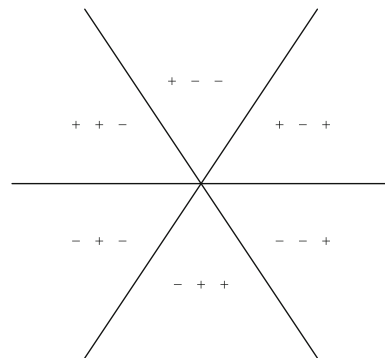
Let M be the centralizer of A in K . Suppose Γ is an irreducible torsion-free cocompact lattice in G . Since A commutes with M , the action of A by right

translations on $\Gamma \backslash G$ descends to an A -action on $N := \Gamma \backslash G / M$. This action is called a *Weyl chamber flow*. Any Weyl chamber flow is an Anosov action, that is, has an element that acts hyperbolically transversally to the orbit foliation of A . Note that all maximal connected $\mathbb{R}$ diagonalizable subgroups of G are conjugate and their common dimension is called the $\mathbb{R}$ -rank of G . If the $\mathbb{R}$ -rank k of G is higher than 2, then the Weyl chamber flow is a higher rank hyperbolic $\mathbb{R}^k$ -action.

An example of semi-simple Lie group is $SL(n, \mathbb{R})$. Let A be the diagonal subgroup of matrices with positive entries in $SL(n, \mathbb{R})$. An example of Weyl chamber flow that will be discussed in the sequel is the action of A by right translations on $\Gamma \backslash SL(n, \mathbb{R})$, where Γ is a cocompact lattice. In this case the centralizer M is trivial. The rank of this action is $n - 1$. The picture of the Weyl chambers for $n = 3$ is shown in Fig. 1. The signs that appear in each chamber are the signs of half of the Lyapunov exponents of a regular element from the chamber with respect to a certain fixed basis. For this action, the Lyapunov exponents appear in pairs of opposite signs.

An example of higher rank lattice Anosov action that will be discussed in the sequel is the standard action of $SL(n, \mathbb{Z})$ on the torus $\mathbb{T}^n$, $(A, x) \mapsto Ax, A \in SL(n, \mathbb{Z}), x \in \mathbb{T}^n$. $SL(n, \mathbb{Z})$ is a (noncocompact!) lattice in $SL(n, \mathbb{R})$. As shown in (Katok and Lewis 1991), this action is generated by Anosov diffeomorphisms.

ERGODIC THEORY: RIGIDITY



Ergodic Theory: Rigidity, Fig. 1 Weyl chambers for $SL(3, \mathbb{R})$

We describe now a class of dynamical systems more general than the hyperbolic one. A C^1 diffeomorphism f of a compact differentiable manifold M is called *partially hyperbolic* if there exists a continuous invariant splitting of the tangent bundle $TM = E^s \oplus E^0 \oplus E^u$ such the derivative of f expands E^u much more than E^0 , and contracts E^s much more than E^0 . See (Brin and Pesin 1974; Burns et al. 2001; Hirsch et al. 1977) for the theory of partially hyperbolic diffeomorphisms. E^s and E^u are called stable, respectively unstable distributions, and are integrable. E^0 is called center distribution and, in general, it is not integrable. A structural stability result proved by Hirsch, Pugh, Shub (Hirsch et al. 1977), that is a frequently used tool in rigidity, shows that, if E^0 is integrable to a smooth foliation, then any perturbation $\bar{f}$ of f is partially hyperbolic and has an integrable center foliation. Moreover, the center foliations of $\bar{f}$ of f are mapped one into the other by a homeomorphism that conjugates the maps induced on the factor spaces of the center foliations by $\bar{f}$ of f respectively.

We review now basic facts about cocycles. These basic definitions refer to several regularity classes: measurable, continuous, or differentiable. Let G be a group acting on a set M , and denote the action $G \times M \rightarrow M$ by $(g, x) \mapsto gx$; thus $(g_1 g_2)x = g_1(g_2 x)$. Let Γ be a group with unit 1_Γ . M is usually endowed with a measurable, continuous, or differentiable structure. A *cocycle* β over the action is a function $\beta : G \times M \rightarrow \Gamma$ such that

$$\beta(g_1 g_2, x) = \beta(g_1, g_2 x) \beta(g_2, x), \quad (1)$$

for all $g_1, g_2 \in G, x \in M$. Note that any group representation $\pi : G \rightarrow \Gamma$ defines a cocycle called *constant cocycle*. The trivial representation defines the *trivial cocycle*.

A natural equivalence relation for the set of cocycles is given by cohomology. Two cocycles $\beta_1, \beta_2 : G \times M \rightarrow \Gamma$ are cohomologous if there exists a map $P : M \rightarrow \Gamma$, called *transfer map*, such that

$$\beta_1(g, x) = P(gx) \beta_2(g, x) P(x)^{-1}, \quad (2)$$

for all $g \in G, x \in M$.

Differentiable Rigidity

We start by reviewing cohomological results. Several basic notions are already defined in section “Basic Definitions and Examples”. In this section we assume that the cocycles are at least continuous. A cocycle $\beta : G \times M \rightarrow \Gamma$ over an action $(g, x) \mapsto gx, g \in G, x \in M$, is said to satisfy *closing conditions* if for any $g \in G$ and $x \in M$ such that $gx = x$, one has $\beta(g, x) = 1_\Gamma$. Note that closing conditions are necessary in order for a cocycle to be cohomologous to the trivial one. Since a $\mathbb{Z}$ -cocycle is determined by a function $\beta : M \rightarrow \Gamma, \beta(x) := \beta(1, x)$, the closing conditions become

$$f^n x = x \quad \text{implies} \quad \beta(f^{n-1}x) \dots \beta(x) = 1_\Gamma, \quad (3)$$

where $f : M \rightarrow M$ is the function that implements the $\mathbb{Z}$ -action.

The first cohomological results for hyperbolic systems were obtained by Livshits (1971, 1972). Let M be a compact Riemannian manifold with a $\mathbb{Z}$ -action implemented by a topologically transitive Anosov C^1 diffeomorphism f . Then an α -Hölder function $\beta : M \rightarrow \mathbb{R}$ determines a cocycle cohomologous to a trivial cocycle if and only if β satisfies the closing conditions (3). The transfer map is α -Hölder. Moreover, for each Hölder class α and each finite dimensional Lie group Γ , there is a neighborhood U of the identity in Γ such that an α -Hölder function $\beta : M \rightarrow \Gamma$ determines a cocycle cohomologous to the trivial cocycle if and only if β satisfies the closing conditions (3). The transfer map is again α -Hölder. Similar results are true for Anosov flows.

Using Fourier analysis, Veech (1986) extended Livshits’s result to real valued cocycles over $\mathbb{Z}$ -actions induced by ergodic endomorphisms of an n -dimensional torus, not necessarily hyperbolic. For cocycles with values in abelian groups, the question of two arbitrary cocycles being cohomologous reduces to the question of an arbitrary cocycle being cohomologous to a trivial one. This is not the case for cocycles with values in non-abelian groups. Parry (1999) extended Livshits’s criteria to one for

cohomology of two arbitrary cocycles with values in compact Lie groups. Parry's result was generalized by Schmidt (1999) to cocycles with values in Lie groups that, in addition, satisfy a center bunching condition. Nițică, Török (1995) extended Livshits's result to cocycles with values in the group $\text{Diff}^k(M)$ of C^k diffeomorphism of a compact manifold M with stably trivial bundle, $k \geq 3$. Examples of such manifolds are the tori and the spheres. In this case, the transfer map takes values in $\text{Diff}^{k-3}(M)$, and it is Hölder with respect to a natural metric on $\text{Diff}^{k-3}(M)$. In (Nițică and Török 1995) one can also find a generalization of Livshits's result to generic Anosov actions, that is, actions generated by families of Anosov diffeomorphisms that do not interchange the stable and unstable directions of elements in the family. An example of such an action is the standard action of $SL(n, \mathbb{Z})$ on the n -dimensional torus.

A question of interest is the following: if two C^k cocycles, $1 \leq k \leq \omega$, over a hyperbolic action, are cohomologous through a continuous/measurable transfer map P , what can be said about the higher regularity of P ? For real valued cocycles the question can be reduced to one about cohomologically trivial cocycles. Livshits showed that for a real valued C^1 cocycle cohomologous to a constant the transfer map is C^1 . He also obtained C^∞ regularity results if the action is given by hyperbolic automorphisms of a torus. After preliminary results by Guillemin and Kazhdan for geodesic flows on surfaces of negative curvature, for general hyperbolic systems the question was answered positively by de la Llave, Marco, Moriyon (1986) in the C^∞ case and by de la Llave (1997) in the real analytic case. Nițică, Török (1998) considered the lift of regularity for a transfer map between two cohomologous cocycles with values in a Lie group or a diffeomorphism group. In contrast to the case of cocycles cohomologous to trivial ones, here one needs to require for the transfer map a certain amount of Hölder regularity that depends on the ratio between the expansion/contraction that appears in the base and the expansion/contraction introduced by the cocycle in the fiber. This

assumption is essential, as follows from a counter example found by de la Llave's (1992).

Useful tools in this development have been results from analysis that lift the regularity of a continuous real valued function which is assumed to have higher regularity along pairs of transverse Hölder foliations. Many times the foliations are the stable and unstable ones associated to a hyperbolic system. Journé (1988) proved the $C^{n,\alpha}$ regularity of a continuous function that is $C^{n,\alpha}$ along two transverse continuous foliations with $C^{n,\alpha}$ leaves. If one is interested only in C^∞ regularity, a convenient alternative is a result of Hurder, Katok (1990). This has a simpler proof and can be applied to the more general situation in which the function is regular along a web of transverse foliations. A real analytic regularity result along these lines belongs to de la Llave (1997). In certain problems, for example when working with Weyl chamber flows, it is difficult to control the regularity in enough directions to span the whole tangent space. Nevertheless, the tangent space can be generated if one consider higher brackets of good directions. A C^∞ regularity result for this case belongs to Katok, Spatzier (1994b). In order to apply this result, the foliations need to be C^∞ not only along the leaves, but also transversally.

An application of the above regularity results is to questions about transverse regularity of the stable and unstable foliations of the geodesic flow on certain C^∞ surfaces of nonpositive curvature. For compact negatively curved C^∞ surfaces, E. Hopf showed that these foliations are C^1 , and it follows from the work of Anosov that individual leaves are C^∞ . Hurder, Katok (1990) showed that once the weak-stable and weak-unstable foliations of a volume-preserving Anosov flow on a compact 3-manifold are C^2 , they are C^∞ .

Another application of the regularity results is to the study of invariants for C^k conjugacy of hyperbolic systems. By structural stability, a small C^1 perturbation of a hyperbolic system is C^0 conjugate to the unperturbed one. The conjugacy, in general, is only Hölder. If the conjugacy is C^1 then it preserves the eigenvalues of the derivative at the periodic orbit. The following two results describe the invariants of smooth

and real analytic conjugacy of low dimensional hyperbolic systems. They are proved in a series of papers written in various combinations by de la Llave, Marco, Moriyo (de la Llave 1987, 1997; de la Llave and Moriyo 1988; Marco and Moriyo 1987a, b).

Let X, Y be two $C^\infty(C^\omega)$ transitive Anosov vector fields on a compact three-dimensional manifold. If they are C^0 conjugate and the eigenvalues of the derivative at the corresponding periodic orbits are the same, then the conjugating homeomorphism is $C^\infty(C^\omega)$. In particular, any C^1 conjugacy is $C^\infty(C^\omega)$.

Assume now that f, g are two $C^\infty(C^\omega)$ Anosov diffeomorphisms on a compact two dimensional manifold. If they are C^0 conjugate and the eigenvalues of the derivative at the corresponding periodic orbits are the same, then the conjugating diffeomorphism is $C^\infty(C^\omega)$. In particular, any C^1 conjugacy is $C^\infty(C^\omega)$.

An important direction was initiated by Katok, Spatzier (1994a) who studied cohomological results over hyperbolic $\mathbb{Z}^k$ or $\mathbb{R}^k$ -actions, $k \geq 2$. They show that real valued smooth/Hölder cocycles over typical classes of hyperbolic $\mathbb{Z}^k$ or $\mathbb{R}^k$, $k \geq 2$, actions are smoothly/Hölder cohomologous to constants. These results cover, in particular, actions by hyperbolic automorphisms of a torus, and Weyl chamber flows. The proofs rely on harmonic analysis techniques, such as Fourier transform and group representations for semi-simple Lie groups.

A geometric method for cocycle rigidity was developed in (Katok et al. 2000). One constructs a differentiable form using invariant structures along stable/unstable foliations, and the commutativity of the action. The form is exact if and only if the cocycle is cohomologous to a constant one. The method covers actions on nilmanifolds satisfying a condition called TNS (TotallyNon-Symplectic). This condition means that the action is higher rank abelian hyperbolic, and that the tangent space is a direct sum of invariant distributions, with each pair of these included in the stable distribution of a hyperbolic element of the action. The method was also applied to small (i.e. close to identity on a set of generators) Lie group valued cocycles. A related paper is (Nițică and Török

2003) which contains rigidity results for cocycles over (TNS) actions with values in compact Lie groups. In this situation the number of cohomology classes is finite. An example of (TNS) action is given by the action of a maximal diagonalizable subgroup of $SL(n, \mathbb{Z})$ on $\mathbb{T}^n$.

Recently Damjanović, Katok (2005) developed a new method that was applied to the action of the matrix diagonal group on $\Gamma \backslash SL(n, \mathbb{R})$. They use techniques from (Katok and Kononenko 1996), where one finds cohomology invariants for cocycles over partially hyperbolic actions that satisfy accessibility property. Accessibility means that one can connect any two points from the manifold supporting the partially hyperbolic dynamical system by transverse piecewise smooth paths included in stable/unstable leaves. This notion was introduced by Brin, Pesin (1974) and it is playing a crucial role in the recent surge of activity in the field of partially hyperbolic diffeomorphisms. See (Burns et al. 2001) for a recent survey of the subject. The cohomology invariants described in (Katok and Kononenko 1996) are heights of the cocycle over cycles constructed in the base out of pieces inside stable/unstable leaves. They provide a complete set of obstructions for solving the cohomology equation. A new tool introduced in (Damjanović and Katok 2005) is algebraic K -theory (Milnor 1971). The method can be extended to cocycles with non-abelian range. In (Katok and Nițică 2007) one finds related results for small cocycles with values in a Lie group or the diffeomorphism group of a compact manifold.

The equivalent of the Livshits theorem in the higher-rank setting appears to be a description of the highest cohomology rather than the first cohomology. Indeed, for higher rank partially hyperbolic actions of the torus, the intermediate cohomologies are trivial, while for the highest one the closing conditions characterize the cohomology classes. This behavior provides a generalization of Veech cohomological result and of Katok, Spatzier cohomological result for toral automorphisms, and was discovered by A. Katok, S. Katok (1995, 2005).

Flaminio, Forni (2003) studied the cohomological equation over the horocycle flow. It is shown

that there are infinitely many obstructions to the existence of a smooth solution. Moreover, if these obstructions vanish, then one can solve the cohomological equation. In (Forni 1997) it is shown a similar result for cocycles over area preserving flows on compact higher-genus surfaces under certain assumptions that hold generically. Mieczkowski (2007) extended these techniques and studied the cohomology of parabolic higher rank abelian actions. All these results rely on noncommutative Fourier analysis, more specifically representation theory of $SL(2, \mathbb{R})$ and $SL(2, \mathbb{C})$.

Local Rigidity

Let Γ be a finitely (or compactly) generated group, G a topological group, and $\pi : \Gamma \rightarrow G$ a homomorphism. The target of local rigidity theory is to understand the space of perturbations of various homomorphisms π . Trivial perturbations of a homomorphism arise from conjugation by an arbitrary element of G . In order to rule them out, one says that π is *locally rigid* if any nearby homomorphism π' , (that is, π' , close to π on a finite or compact set of generators of Γ), is conjugate to π by an element $g \in G$, that is, $\pi(\gamma) = g\pi'(\gamma)g^{-1}$ for all $\gamma \in \Gamma$. If G is path-wise connected, one can also consider *deformation rigidity*, meaning that any nearby continuous path of homomorphisms is conjugated to the initial one via a continuous path of elements in G that has an end in the identity.

Initial results on local rigidity are about embeddings of lattices into semi-simple Lie groups. The main results belong to Weil (1960, 1962, 1964). He showed that if G is a semi-simple Lie group that is not locally isomorphic to $SL(2, \mathbb{R})$ and if $\Gamma \subset G$ is an irreducible cocompact lattice, then the natural embedding of Γ into G is locally rigid. Earlier results were obtained by Selberg (1960), Calabi, Vesentini (1960), and Calabi (1961). Selberg proved the local rigidity of the natural embedding of cocompact lattices into $SL(n, \mathbb{R})$. His proof used the dynamics of iterates of matrices, in particular the existence of singular directions, or walls of Weyl chambers, in the maximal diagonalizable subgroups of $SL(n, \mathbb{R})$. Selberg's

approach inspired Mostow (1973) to use the boundaries at infinity in his proof of strong rigidity of lattices, which in turn was crucial to the development of superrigidity due to Margulis (1991). See section “Global Rigidity” for more details.

Recall that hyperbolic dynamical systems are structurally stable. Thus they are, in a certain sense, locally rigid. We introduce now a precise definition of local rigidity in the infinite-dimensional setup. The fact that for general group actions one needs to consider different regularities for the actions, perturbations and conjugacies is apparent from the description of structural stability for Anosov systems.

A C^k action α of a finitely generated discrete group Γ on a manifold M , that is, a homomorphism $\alpha : \Gamma \rightarrow \text{Diff}^k(M)$, is said to be $C^{k,l,p,r}$ *locally rigid* if any C^l perturbation $\tilde{\alpha}$ which is C^p close to α on a family of generators, is C^r conjugate to α , i.e. there exists a C^r diffeomorphism $H : M \rightarrow M$ which conjugates $\tilde{\alpha}$ to α , that is, $H \circ \alpha(g) = \tilde{\alpha}(g) \circ H$ for all $g \in \Gamma$. Note that for Anosov $\mathbb{Z}$ -actions, $C^{1,1,1,0}$ rigidity is known as *structural stability*. One can also introduce the notion of *deformation rigidity* if the initial action and the perturbation are conjugate by a continuous path of diffeomorphisms that has an end coinciding to the identity.

A weaker notion of local rigidity can be defined in the presence of invariant foliations for the initial group action and for the perturbation. The map H is now required to preserve the leaves of the foliations and to conjugate only after factorization by the invariant foliations. The importance of this notion is apparent from the leaf wise conjugacy structural stability theorem of Hirsch, Pugh, Shub (1977). See section “Basic Definitions and Examples”. Moreover, for Anosov flows this is the natural notion of structural stability, and appears by taking the invariant foliation to be the one-dimensional orbit foliation. For more general actions, of lattices or higher rank abelian groups, this property is often used in combination to cocycle rigidity in order to show local rigidity. We discuss more about this when we review local rigidity results for partially hyperbolic actions.

We summarize now several developments in local rigidity that emerged in the 80s. Initial results (Lewis 1991; Zimmer 1990) were about *infinitesimal rigidity*, that is, a weaker version of local rigidity suitable for discrete groups representations in infinite-dimensional spaces of smooth vector fields. Then Hurder (1992) proved $C^{\infty, \infty, 1, \infty}$ deformation rigidity and Katok, Lewis, Zimmer (Katok and Lewis 1991, 1996; Katok et al. 1996) proved $C^{\infty, \infty, 1, \infty}$ local rigidity of the standard action of $SL(n, \mathbb{Z})$, $n \geq 3$, on the n -dimensional torus. In these results crucial use was made of the presence of an Anosov element in the action. Due to the uniqueness of the conjugacy coming from structural stability, one has a continuous candidate for the conjugacy between the actions. Margulis, Qian (2001) used the existence of a spanning family of directions that are hyperbolic for certain elements of the action to show local rigidity of partially hyperbolic actions that are not hyperbolic. Another important tool present in many proofs is Margulis and Zimmer super-rigidity. These results allow one to produce a measurable conjugacy for the perturbation. Then one shows that the conjugacy has higher regularity using the presence of hyperbolicity. Having enough directions to span the whole tangent space is essential to lift the regularity. A cocycle to which superrigidity can be applied is the derivative cocycle.

The study of local rigidity of partially hyperbolic actions that contain a compact factor was initiated by Nițică, Török (1995, 2001). Let $n \geq 3$ and $d \geq 1$. Let ρ be the action of $SL(n, \mathbb{Z})$ on $\mathbb{T}^{n+d} = \mathbb{T}^n \times \mathbb{T}^d$ given by $\rho(A)(x, y) = (Ax, y)$, $x \in \mathbb{T}^n$, $y \in \mathbb{T}^d$, $A \in SL(n, \mathbb{Z})$. Then, for $K \geq 1$, (Nițică and Török 1995) shows that ρ is $C^{\infty, \infty, 5, K-1}$ deformation rigid. The proof is based on three results in hyperbolic dynamics: the generalization of Livshits's cohomological results to cocycles with values in diffeomorphism groups, the extension of Livshits's result to general Anosov actions, and a version of the Hirsch, Pugh, Shub structural stability theorem improving the regularity of the conjugacy.

Assume now $n \geq 3$ and $K \geq 1$. If ρ is the action of $SL(n, \mathbb{Z})$ on $\mathbb{T}^{n+1} = \mathbb{T}^n \times \mathbb{T}$ given by $\rho(A)(x, y) = (Ax, y)$, $x \in \mathbb{T}^n$, $y \in \mathbb{T}$, $A \in SL(n, \mathbb{Z})$,

(Nițică and Török 2001) shows that the action ρ is $C^{\infty, \infty, 2, K-1}$ locally rigid. Ingredients in the proof are two rigidity results, one about TNS actions, and one about actions of property (T) groups. A locally compact group has property (T) if the trivial representation is isolated in the Fell topology. This means that if G acts on a Hilbert space unitarily and it has almost invariant vectors, then it has invariant vectors. Hirsch–Pugh–Shub theorem implies that perturbations of abelian partially hyperbolic actions of product type are conjugated to skew-products of abelian Anosov actions via cocycles with values in diffeomorphism groups. In addition, the TNS property implies that the sum of the stable and unstable distributions of any regular element of the perturbation is integrable. The leaves of the integral foliation are closed, covering the base simply. Thus one obtains a conjugacy between the perturbation and a product action. Property (T) is used to show that the conjugacy reduces the perturbed action to a family of perturbations of hyperbolic actions. But the last ones are already known to be conjugate to the hyperbolic action in the base.

Recent important progress in the question of local rigidity of lattice actions was made by Fisher, Margulis (2003, 2004, 2005). Their proofs are modeled along the proof of Weil's local rigidity result (Weil 1964) and use an analog of Hamilton's (1982) hard implicit function theorem. Let G be a connected semi-simple Lie group with all simple factors of rank at least two, and $\Gamma \subset G$ a lattice. The main result shows that a volume preserving affine action ρ of G or Γ on a compact smooth manifold X is $C^{\infty, \infty, \infty, \infty}$ locally rigid. Lower regularity results are also available. A component of the proof shows that if Γ is a group with property (T), X a compact smooth manifold, and ρ a smooth action of Γ on X by Riemannian isometries, then ρ is $C^{\infty, \infty, \infty, \infty}$ locally rigid. An earlier local rigidity result for this type of actions by cocompact lattices was obtained by Benveniste (2000).

Many lattices act naturally on “boundaries” of type G/P , where G is a semi-simple algebraic Lie group and P is a parabolic subgroup. An example is given by $G = SL(2, \mathbb{R})$ and P the subgroup in G consisting of upper triangular matrices. Local

rigidity results for this type of actions were found by Ghys (1985), Kanai (1996) and Katok, Spatzier (1997).

Starting with the work of Katok and Lewis, a related direction was the study of local rigidity for higher rank abelian actions. They prove in (Katok and Lewis 1991) the $C^{\infty,\infty,1,\infty}$ local rigidity of the action of a $\mathbb{Z}^n$ maximal diagonalizable (over $\mathbb{R}$) subgroup of $SL(n+1, \mathbb{Z})$, $n \geq 2$, acting on the torus $\mathbb{T}^{n+1}$. These type of results were later pushed forward by Katok, Spatzier (1997). Using the theory of nonstationary normal forms developed in (Guysinsky 2002; Guysinsky and Katok 1998) by Katok, Guysinsky, they proved several local rigidity results. The first one assumes that G is a semi simple Lie group with all simple factors of rank at least two, Γ a lattice in G , N a nilpotent Lie group and Λ a lattice in N . Then any Anosov affine action of Γ on N/Λ is $C^{\infty,\infty,1,\infty}$ locally rigid. Second, let $\mathbb{Z}^d$ be a group of affine transformations of N/Λ for which the derivatives are simultaneously diagonalizable over $\mathbb{R}$ with no eigenvalues on the unit circle. Then the $\mathbb{Z}^d$ -action on N/Λ is $C^{\infty,\infty,1,\infty}$ locally rigid. A related result for continuous groups is the $C^{\infty,\infty,1,\infty}$ local rigidity (after factorization by the orbit foliation) of the action of a maximal abelian $\mathbb{R}$ -split subgroup in an $\mathbb{R}$ -split semi-simple Lie group of real rank at least two on G/Λ , where Λ is a cocompact lattice in G .

One can also study rigidity of higher rank abelian partially hyperbolic actions that are not hyperbolic. Natural examples appear as automorphisms of tori and as variants of Weyl chamber flows. For the case of ergodic actions by automorphisms of a torus, this was investigated using a version of the KAM (Kolmogorov, Arnold, Moser) method by Damianović, Katok (2007). As usual in the KAM method, one starts with a linearization of the conjugacy equation. At each step of the iterative KAM scheme, some twisted cohomological equations are solved. The existence of the solutions is forced by the ergodicity of the action and the higher rank assumptions. Diophantine conditions present in this case allow to control the fixed loss of regularity which is necessary for the convergence of these solutions to a conjugacy.

Global Rigidity

The first remarkable result in global rigidity belongs to Mostow (1973). For G a connected non-compact semi-simple Lie group not locally isomorphic to $SL(2, \mathbb{R})$, and two irreducible cocompact lattices $\Gamma_1, \Gamma_2 \subset G$, Mostow showed that any isomorphism θ from Γ_1 into Γ_2 extends to an isomorphism of G into itself. G has an involution σ whose fixed set is a maximal compact subgroup K . One constructs the symmetric Riemannian space $X = G/K$. To each chamber of X corresponds a parabolic group and these parabolic groups are endowed with a Tits geometry similar to the projective geometry of lines, planes etc. formed in the classical case when $G = PGL(n, \mathbb{R})$. The proof of Mostow's result starts by building a θ -equivariant pseudo-isometric map $\phi : G/K_1 \rightarrow G/K_2$. The map ϕ induces an incidence preserving θ -equivariant isomorphism ϕ_0 of the Tits geometries. By Tits' generalized fundamental theorem of projective geometry, ϕ_0 is induced by an isomorphism of G . Finally, $\theta(\gamma) = \phi_0 \cdot \gamma \cdot \phi_0^{-1}$ gives the desired conclusion.

The next remarkable result in global rigidity is Margulis' superrigidity theorem. An account of this development can be found in the monograph (Margulis 1991). For large classes of irreducible lattices in semi-simple Lie groups, this result classifies all finite dimensional representations. Let G be a semi-simple simply connected Lie group of rank higher than two and $\Gamma < G$ an irreducible lattice. Then any linear representation π of Γ is almost the restriction of a linear representation of G . That is, there exists a linear representation π_1 of G and a bounded image representation π_2 of Γ such that $\pi = \pi_1 \pi_2$. The possible representations π_2 are also classified by Margulis up to some facts concerning finite image representations. As in the case of Mostow's result, the proof involved the analysis of a map defined on the boundary at infinity. In this case the map is studied using deep results from dynamics like the multiplicativity ergodic theorem of Oseledec (1968) or the theory of random walks on groups developed by Furstenberg (1963). An important consequence of Margulis superrigidity result is the

arithmeticity of irreducible lattices in connected semi-simple Lie groups of rank higher than two. A basic example of arithmetic lattice can be obtained by taking the integer points in a semi-simple Lie group that is a matrix group, like taking $SL(n, \mathbb{Z})$ inside $SL(n, \mathbb{R})$. Special cases of superrigidity theorems were proved by Corlette (1992) and Gromov, Schoen (1992) for the rank one groups $Sp(1, n)$ and respectively F_4 using the theory of harmonic maps. A consequence is the arithmeticity of lattices in these groups. Some of these results are put into differential geometric setting in (Mok et al. 1993).

Margulis superrigidity result was extended to cocycles by Zimmer. A detailed exposition, including a self contained presentation of several rigidity results of Margulis, can be found in the monograph (Zimmer 1984). We mention here a version of this result that can be found in (Fisher and Margulis 2003). Let M be a compact manifold, H a matrix group, $P = M \times H$, and Γ a lattice in a simply connected, semi-simple Lie group with all factors of rank higher than two. Assume that Γ acts on M and H in a way that makes the projection from P to M equivariant. Moreover, the action of Γ on P is measure preserving and ergodic. Then there exists a measurable map $s : M \rightarrow H$, a representation $\pi : G \rightarrow H$, a compact subgroup $K < H$ which commute with $\pi(G)$ and a measurable map $k : \Gamma \times M \rightarrow K$ such that $\gamma \cdot s(m) = k(\gamma, m) \cdot \pi(\gamma) \cdot s(\gamma \cdot m)$. One can easily check from the last equation that k is a cocycle. So, up to a measurable change of coordinates given by the map s , the action of Γ on P is a compact extension via a cocycle of a linear representation of G .

Developing further the method of Mostow for studying the Tits building associated to a symmetric space of non-positive curvature led Ballman, Brin, Eberlein, Spatzier (1985a, b) to a number of characterizations of symmetric spaces. In particular, they showed that if M is a complete Riemannian manifold of non-positive curvature, finite volume, with simply connected cover, irreducible and of rank at least two, then M is isometric to a symmetric space with the connected component of $\text{Isom}(M)$ having no compact factors.

A topological rigidity theorem has been proved by Farrell, Jones (1989). They showed that if N is a complete connected Riemannian manifold whose sectional curvature lies in a closed interval included in $(-\infty, 0]$, and M is a topological manifold of dimension greater than 5, then any proper homotopy equivalence $f : M \rightarrow N$ is properly homotopic to a homeomorphism. In particular, if M and N are both compact connected negatively curved Riemannian manifolds with isomorphic fundamental groups, then M and N are homeomorphic.

Likewise to the case of local rigidity, a source of inspiration for results in global rigidity was the theory of hyperbolic systems, in particular their classification. The only known examples of Anosov diffeomorphisms are hyperbolic automorphisms of infranilmanifolds. Moreover, any Anosov diffeomorphism on an infranilmanifold is topologically conjugate to a hyperbolic automorphism (Franks 1970; Manning 1974). It is conjectured that any Anosov diffeomorphism is topologically conjugate to a hyperbolic automorphism of an infranilmanifold. Partial results are obtained in (Newhouse 1970), where the conjecture is proved for Anosov diffeomorphisms with codimension one stable/unstable foliation. The proof of the general conjecture eluded efforts done so far. It is not even known if any Anosov diffeomorphism is topologically transitive, that is, if it has a dense orbit. A few positive results are available. Let M be a C^∞ compact manifold endowed with a C^∞ affine connection. Let f be a topologically transitive Anosov diffeomorphism preserving the connection and such that the stable and unstable distributions are C^∞ . Then Benoist, Labourie (1993) proved that f is C^∞ conjugate to a hyperbolic automorphism of an infranilmanifold.

The situation for Anosov flows is somehow different. As shown in (Franks and Williams 1980), there exist Anosov flows that are not topologically transitive, so a general analog of the conjecture is false. Nevertheless, for the case of codimension one stable or unstable foliation, it is conjectured in (Verjovsky 1974) that any Anosov flow on a manifold of dimension greater than three admits a global cross-section. This would

imply that the flow is topologically conjugate to the suspension of a linear automorphism of a torus.

For actions of groups larger than $\mathbb{Z}$, or $\mathbb{R}$, global classification results are more abundant. A useful strategy in these type of results, which are quite technical, is to start by obtaining a measurable description of the action, most of the time using Margulis–Zimmer superrigidity results, and then use extra assumptions on the action, such as the presence of a hyperbolic element, or the presence of an invariant geometric structure, or both, in order to show that the measurable model is actually continuous or even differentiable. For actions of higher rank Lie groups and their lattices some representative papers are by Katok, Lewis, Zimmer (1996) and Goetze, Spatzier (1999). For actions of higher rank abelian groups see Kalinin, Spatzier (2007).

Measure Rigidity

Measure rigidity is the study of invariants measures for actions of one parameter and multi parameter abelian groups and semigroups acting on manifolds. Typical situations when interesting rigidity phenomena appear are for one parameter unipotent actions and higher rank hyperbolic actions, discrete or continuous.

A unipotent matrix is one all of whose eigenvalues are one. An important case where the action of a unipotent flow appears is that of the horocycle flow. The invariant measures for it were studied by Furstenberg (1973), who showed that the horocycle flow on a compact surface is uniquely ergodic, that is, the ergodic measure is unique. Dani and Smillie (1984) extended this result to the case of non-compact surfaces, with the only other ergodic measures appearing being those supported on compact horocycles. An important breakthrough is the work of Margulis (1989), who solved a long standing question in number theory, Oppenheim’s conjecture. The conjecture is about the density properties of the values of an indefinite quadratic form in three or more variables, provided the form is not

proportional to a rational form. The proof of the conjecture is based on the study of the orbits for unipotent flows acting by translations on the homogenous space $SL(n, \mathbb{Z}) \backslash SL(n, \mathbb{R})$. All these results were special cases of the Raghunathan conjecture about the structure of the orbits of the actions of unipotent flows on homogenous spaces. Raghunathan’s conjecture was proved in full generality by Ratner (1991a, b). Borel, Prasad (1992) raised the question of an analog of Raghunathan’s conjecture for S -algebraic groups. S -algebraic groups are products of real and p -adic algebraic groups. This was answered independently by Margulis, Tomanov (1994) and Ratner (1995).

A basic example of higher rank abelian hyperbolic action is given by the action of $S_{m,n}$, the multiplicative semigroup of endomorphisms generated by the multiplication by m and n , two nontrivial integers, on the one dimensional torus $\mathbb{T}^1$. In a pioneering paper (Furstenberg 1967) Furstenberg showed that for m, n that are not powers of the same integer the action of $S_{m,n}$ has a unique closed, infinite invariant set, namely $\mathbb{T}^1$ itself. Since there are many closed, infinite invariant sets for multiplication by m , and by n , this result shows a remarkable rigidity property of the joint action. Furstenberg’s result was generalized later by Berend for other group actions on higher dimensional tori and on other compact abelian groups in (Berend 1983, 1984).

Furstenberg further opened the field by raising the following question:

Conjecture 1 *Let μ be a $S_{m,n}$ -invariant and ergodic probability measure on $\mathbb{T}^1$. Then μ is either an atomic measure supported on a finite union of (rational) periodic orbits or μ is the Lebesgue measure.*

While the statement appears to be simple, proving it has been elusive. The first partial result was given by Lyons (1988) under the strong additional assumption that the measure makes one of the endomorphisms generating the action exact. Later Rudolph (1990) and Johnson (1992) weaken the exactness assumption and proved

that μ must be the Lebesgue measure provided that multiplication by m (or multiplication by n) has positive entropy with respect to μ . Their results have been proved again using slightly different methods by Feldman (1993). A further extension was given by Host (1995).

Katok proposed another example for which measure rigidity can be fruitfully tested, the $\mathbb{Z}^2$ -action induced by two commuting hyperbolic automorphisms on the torus $\mathbb{T}^3$. An example of such action is shown in section “Basic Definitions and Examples”. One can also consider the action induced by a $\mathbb{Z}^{n-1}$ maximal abelian group of hyperbolic automorphisms acting on the torus $\mathbb{T}^n$. Katok, Spatzier (1996) developed a more geometric technique allowing to prove measure rigidity for these actions if they have an element acting with positive entropy. Their technique can be applied in higher generality if the action is irreducible in a strong sense and, in addition, it has individual ergodic elements or it is TNS. See also (Kalinin and Katok 2002). This method is based on the study of conditional measures induced by the invariant measure on certain invariant foliations that appear naturally in the presence of a hyperbolic action. Besides stable and unstable foliations, one can also consider various intersections of them. Einsiedler, Lindenstrauss (2003) were able to eliminate the ergodicity and TNS assumptions.

Yet another interesting example of higher rank abelian action is given by Weyl chamber flows. These do not satisfy the TNS condition. Einsiedler, Katok (2003) proved that if G is $SL(n, \mathbb{R})$, $\Gamma \subset G$ is a lattice, H is the subgroup of positive diagonal matrices in G , and μ a H -invariant and ergodic measure on G/Γ such that the entropy of μ with respect to all one parameter subgroups of H is positive, then μ is the G invariant measure on G/Γ .

These results are useful in the investigation of several deep questions in number theory. Define $X = SL(3, \mathbb{Z}) \backslash SL(3, \mathbb{R})$ the diagonal subgroup of matrices with positive entries in $SL(3, \mathbb{R})$. This space is not compact but is endowed with a unique translation invariant probability measure. The diagonal subgroup

$$H = \left\{ \begin{pmatrix} e^s & 0 & 0 \\ 0 & e^t & 0 \\ 0 & 0 & e^{-s-t} \end{pmatrix} : s, t \in \mathbb{R} \right\}$$

acts naturally on X by right translations. It was conjectured by Margulis (1997) that any compact H -invariant subset of X is a union of compact H -orbits. A positive solution to this conjecture implies a long standing conjecture of Littlewood:

Conjecture 2 Let $\|x\|$ denote the distance between the real number x and the closest integer. Then

$$\liminf_{n \rightarrow \infty} n \| \alpha \| \| n \beta \| = 0 \quad (4)$$

for any real numbers α and β .

A partial result was obtained by Einsiedler, Katok, Lindenstrauss (2006) who proved that the set of pairs $(\alpha, \beta) \in \mathbb{R}^2$ for which (4) is not satisfied has Hausdorff dimension zero. Applications of these techniques to questions in quantum ergodicity were found by Lindenstrauss (2006).

A current direction in measure rigidity is attempting to classify the invariant measures under rather weak assumptions about the higher rank abelian action, like the homotopical data for the action. Kalinin and Katok (2007) proved that any $\mathbb{Z}^k$, $k \geq 2$, smooth action α on a $k + 1$ -dimensional torus whose elements are homotopic to the corresponding elements of an action by hyperbolic automorphisms preserves an absolutely continuous measure.

Future Directions

An important open problem in differential rigidity is to find invariants for the C^k conjugacy of the perturbations of higher dimensional hyperbolic systems. For Anosov diffeomorphisms, de la Llave counter example (de la Llave 1992) shows that this extension is not possible for a four dimensional example that appears as a direct product of two dimensional Anosov diffeomorphisms.

Indeed, there are C^∞ perturbations of the product that are only C^k conjugate to the unperturbed system for any $k \geq 1$. In the positive direction, Katok conjectured that generalizations are possible for the diffeomorphism induced by an irreducible hyperbolic automorphism of a torus. One can also investigate this question for Anosov flows.

Examples of higher rank Lie groups can be obtained by taking products of rank one Lie groups. Many actions of irreducible lattices in these type of groups are believed to be locally rigid, but the techniques available so far cannot be applied. A related problem is to study local rigidity in low regularity classes, for example the local rigidity of homomorphisms from higher rank lattices into homeomorphism groups. More problems related to local rigidity can be found in (Fisher 2006).

An important problem in global rigidity, emphasized by Katok and Spatzier, is to show that, up to differentiable conjugacy, any higher rank Anosov or partially hyperbolic action is algebraic under the assumption that it is sufficiently irreducible. The irreducibility assumption is needed in order to exclude actions obtained by successive application of products, extensions, restrictions and time changes from basic ingredients which include some rank one actions.

Another problem of current research in measure rigidity is to develop a counterpart of Ratner's theory for the case of actions by hyperbolic higher rank abelian groups on homogeneous spaces. It was conjectured by Katok, Spatzier (1996) that the invariant measures for such actions given by toral automorphisms or Weyl chamber flows are essentially algebraic, that is, supported on closed orbits of connected subgroups. Margulis in (Margulis 2000) extended this conjecture to a rather general setup addressing both the topological and measurable aspects of the problem. More details about actions on homogeneous spaces, as well as connections to diophantine approximation, can be found in the article ► [Ergodic Theory on Homogeneous Spaces and Metric Number Theory](#) by Kleinbock.

Acknowledgment This research was supported in part by NSF Grant DMS-0500832.

Bibliography

Primary Literature

- Anosov DV (1967) Geodesic flows on closed Riemannian manifolds with negative curvature. *Proc Stek Inst* 90: 1–235
- Ballman W, Brin M, Eberlein P (1985a) Structure of manifolds of non-negative curvature. I. *Ann Math* 122: 171–203
- Ballman W, Brin M, Spatzier R (1985b) Structure of manifolds of non-negative curvature. II. *Ann Math* 122: 205–235
- Benoist Y, Labourie F (1993) Flots d'Anosov à distributions stable et instable différentiables. *Invent Math* 111:285–308
- Benveniste EJ (2000) Rigidity of isometric lattice actions on compact Riemannian manifolds. *Geom Funct Anal* 10:516–542
- Berend D (1983) Multi-invariant sets on tori. *Trans AMS* 280:509–532
- Berend D (1984) Multi-invariant sets on compact abelian groups. *Trans AMS* 286:505–535
- Borel A, Prasad G (1992) Values of isotropic quadratic forms at S-integral points. *Compos Math* 83: 347–372
- Brin MI, Pesin YA (1974) Partially hyperbolic dynamical systems. *Izvestia* 38:170–212
- Burns K, Pugh C, Shub M, Wilkinson A (2001) Recent results about stable ergodicity. In: Katok A, Pesin Y, de la Llave R, Weiss H (eds) *Smooth ergodic theory and its applications*, Seattle, 1999. *Proc Symp Pure Math*, vol 69. AMS, Providence, pp 327–366
- Calabi E (1961) On compact Riemannian manifolds with constant curvature. I. *Proc Symp Pure Math*, vol 3. AMS, Providence, pp 155–180
- Calabi E, Vesentini E (1960) On compact, locally symmetric Kähler manifolds. *Ann Math* 17:472–507
- Corlette K (1992) Archimedean superrigidity and hyperbolic geometry. *Ann Math* 135:165–182
- Damjanović D, Katok A (2005) Periodic cycle functionals and cocycle rigidity for certain partially hyperbolic actions. *Discrete Contin Dynam Syst* 13: 985–1005
- Damjanović D, Katok A (2007) Local rigidity of partially hyperbolic actions KAM, I. Method and Z^k -actions on the torus. Preprint available at http://www.math.psu.edu/katok_a
- Dani SG, Margulis GA (1990) Values of quadratic forms at integer points: an elementary approach. *Enseign Math* 36:143–174
- Dani SG, Smillie J (1984) Uniform distributions of horocycle orbits for Fuchsian groups. *Duke Math J* 51:185–194
- de la Llave R (1987) Invariants for smooth conjugacy of hyperbolic dynamical systems. I. *Commun Math Phys* 109:369–378
- de la Llave R (1992) Smooth conjugacy and S-R-B measures for uniformly and non-uniformly hyperbolic dynamical systems. *Commun Math Phys* 150:289–320

- de la Llave R (1997) Analytic regularity of solutions of Livshits's cohomology equation and some applications to analytic conjugacy of hyperbolic dynamical systems. *Ergodic Theory Dynam Syst* 17:649–662
- de la Llave R, Moriyon R (1988) Invariants for smooth conjugacy of hyperbolic dynamical systems. IV. *Commun Math Phys* 116:185–192
- de la Llave R, Marco JM, Moriyon R (1986) Canonical perturbation theory of Anosov systems and regularity results for the Livšic cohomology equation. *Ann Math* 123:537–611
- Einsiedler M, Katok A (2003) Invariant measures on G/T for split simple Lie groups. *Commun Pure Appl Math* 56:1184–1221
- Einsiedler M, Lindenstrauss E (2003) Rigidity properties of $\mathbb{Z}^d$ -actions on tori and solenoids. *Electron Res Announc AMS* 9:99–110
- Einsiedler M, Katok A, Lindenstrauss E (2006) Invariant measures and the set of exceptions to Littlewood conjecture. *Ann Math* 164:513–560
- Farrell FT, Jones LE (1989) A topological analog of Mostow's rigidity theorem. *J AMS* 2:237–370
- Feldman J (1993) A generalization of a result of R Lyons about measures on $[0,1]$. *Israel J Math* 81:281–287
- Fisher D (2006) Local rigidity of group actions: past, present, future. In: *Dynamics, ergodic theory and geometry* (2007). Cambridge University Press
- Fisher D, Margulis GA (2003) Local rigidity for cocycles. In: *Surv Diff Geom VIII*. International Press, Cambridge, pp 191–234
- Fisher D, Margulis GA (2004) Local rigidity of affine actions of higher rank Lie groups and their lattices. 2003
- Fisher D, Margulis GA (2005) Almost isometric actions, property T, and local rigidity. *Invent Math* 162:19–80
- Flaminio L, Forni G (2003) Invariant distributions and time averages for horocycle flows. *Duke Math J* 119:465–526
- Forni G (1997) Solutions of the cohomological equation for area-preserving flows on compact surfaces of higher genus. *Ann Math* 146:295–344
- Franks J (1970) Anosov diffeomorphisms. In: Chern SS, Smale S (eds) *Global analysis (Proc Symp Pure Math, XIV, Berkeley 1968)*. AMS, Providence, pp 61–93
- Franks J, Williams R (1980) Anomalous anosov flows. In: *Global theory of dynamical systems, Proc Inter Conf Evanston, 1979*. Lecture notes in mathematics, vol 819. Springer, Berlin, pp 158–174
- Furstenberg H (1963) A Poisson formula for semi-simple Lie groups. *Ann Math* 77:335–386
- Furstenberg H (1967) Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math Syst Theory* 1:1–49
- Furstenberg H (1973) The unique ergodicity of the horocycle flow. In: *Recent advances in topological dynamics, Proc Conf Yale Univ, New Haven 1972*. Lecture notes in mathematics, vol 318. Springer, Berlin, pp 95–115
- Ghys E (1985) Actions localement libres du groupe affine. *Invent Math* 82:479–526
- Goetze E, Spatzier R (1999) Smooth classification of Cartan actions of higher rank semi-simple Lie groups and their lattices. *Ann Math* 150:743–773
- Gromov M, Schoen R (1992) Harmonic maps into singular spaces and p -adic superrigidity for lattices in groups of rank one. *Publ Math IHES* 76:165–246
- Guysinsky M (2002) The theory of nonstationary normal forms. *Ergodic Theory Dynam Syst* 22:845–862
- Guysinsky M, Katok A (1998) Normal forms and invariant geometric structures for dynamical systems with invariant contracting foliations. *Math Res Lett* 5:149–163
- Hamilton R (1982) The inverse function theorem of Nash and Moser. *Bull AMS* 7:65–222
- Helgason S (1978) *Differential geometry, Lie groups and symmetric spaces*. Academic, New York
- Hirsch M, Pugh C, Shub M (1977) *Invariant manifolds*. Lecture notes in mathematics, vol 583. Springer, Berlin
- Host B (1995) *Nombres normaux, entropie, translations*. *Israel J Math* 91:419–428
- Hurder S (1992) Rigidity of Anosov actions of higher rank lattices. *Ann Math* 135:361–410
- Hurder S, Katok A (1990) Differentiability, rigidity and Godbillon-Vey classes for Anosov flows. *Publ Math IHES* 72:5–61
- Johnson AS (1992) Measures on the circle invariant under multiplication by a nonlacunary subsemigroup of integers. *Israel J Math* 77:211–240
- Journé JL (1988) A regularity lemma for functions of several variables. *Rev Mat Iberoam* 4:187–193
- Kalinin B, Katok A (2002) Measurable rigidity and disjointness for $\mathbb{Z}^k$ -actions by toral automorphisms. *Ergodic Theory Dynam Syst* 22:507–523
- Kalinin B, Katok A (2007) Measure rigidity beyond uniform hyperbolicity: invariant measures for Cartan actions on tori. *J Modern Dyn* 1:123–146
- Kalinin B, Spatzier R (2007) On the classification of Cartan actions. *Geom Funct Anal* 17:468–490
- Kanai M (1996) A new approach to the rigidity of discrete group actions. *Geom Funct Anal* 6:943–1056
- Katok A, Hasselblatt B (1995) *Introduction to the modern theory of dynamical systems*. Encyclopedia of mathematics and its applications, vol 54. Cambridge University Press, Cambridge
- Katok A, Katok S (1995) Higher cohomology for abelian groups of toral automorphisms. *Ergodic Theory Dynam Syst* 15:569–592
- Katok A, Katok S (2005) Higher cohomology for abelian groups of toral automorphisms. II. The partially hyperbolic case, and corrigendum. *Ergodic Theory Dynam Syst* 25:1909–1917
- Katok A, Kononenko A (1996) Cocycles' stability for partially hyperbolic systems. *Math Res Lett* 3:191–210
- Katok A, Lewis J (1991) Local rigidity for certain groups of toral automorphisms. *Israel J Math* 75:203–241

- Katok A, Lewis J (1996) Global rigidity results for lattice actions on tori and new examples of vol preserving actions. *Israel J Math* 93:253–280
- Katok A, Nițică V (2007) Rigidity of higher rank abelian cocycles with values in diffeomorphism groups. *Geom Dedicata* 124:109–131
- Katok A, Nițică V. Differentiable rigidity of abelian group actions. Cambridge University Press (to appear)
- Katok A, Spatzier R (1994a) First cohomology of Anosov actions of higher rank abelian groups and applications to rigidity. *Publ Math IHES* 79:131–156
- Katok A, Spatzier R (1994b) Subelliptic estimates of polynomial differential operators and applications to rigidity of abelian actions. *Math Res Lett* 1:193–202
- Katok A, Spatzier R (1996) Invariant measures for higher-rank abelian actions. *Ergodic Theory Dynam Syst* 16:751–778; Katok A, Spatzier R (1998) Corrections to: invariant measures for higher-rank abelian actions.; (1996) *Ergodic Theory Dynam Syst* 16:751–778; *Ergodic Theory Dynam Syst* 18:503–507
- Katok A, Spatzier R (1997) Differential rigidity of Anosov actions of higher rank abelian groups and algebraic lattice actions. *Trudy Mat Inst Stek* 216:292–319
- Katok A, Lewis J, Zimmer R (1996) Cocycle superrigidity and rigidity for lattice actions on tori. *Topology* 35:27–38
- Katok A, Nițică V, Török A (2000) Non-abelian cohomology of abelian Anosov actions. *Ergodic Theory Dynam Syst* 2:259–288
- Katok A, Katok S, Schmidt K (2002) Rigidity of measurable structure for $\mathbb{Z}^d$ -actions by automorphisms of a torus. *Comment Math Helv* 77:718–745
- Kazhdan DA (1967) On the connection of a dual space of a group with the structure of its closed subgroups. *Funkc Anal Prilozen* 1:71–74
- Lewis J (1991) Infinitesimal rigidity for the action of $SL_n(\mathbb{Z})$ on T^n . *Trans AMS* 324:421–445
- Lindenstrauss E (2005) Rigidity of multiparameter actions. *Israel Math J* 149:199–226
- Lindenstrauss E (2006) Invariant measures and arithmetic quantum unique ergodicity. *Ann Math* 163:165–219
- Livshits A (1971) Homology properties of Y -systems. *Math Zametki* 10:758–763
- Livshits A (1972) Cohomology of dynamical systems. *Izvestia* 6:1278–1301 the Livšic cohomology equation. *Ann Math* 123:537–611
- Lyons R (1988) On measures simultaneously 2- and 3-invariant. *Israel J Math* 61:219–224
- Manning A (1974) There are no new Anosov diffeomorphisms on tori. *Am J Math* 96:422–429
- Marco JM, Moriyo R (1987a) Invariants for smooth conjugacy of hyperbolic dynamical systems. I. *Commun Math Phys* 109:681–689
- Marco JM, Moriyo R (1987b) Invariants for smooth conjugacy of hyperbolic dynamical systems. III. *Commun Math Phys* 112:317–333
- Margulis GA (1989) Discrete subgroups and ergodic theory. In: *Number theory, trace formulas and discrete groups*, Oslo, 1987. Academic, Boston, pp 277–298
- Margulis GA (1991) Discrete subgroups of semi-simple Lie groups. Springer, Berlin
- Margulis GA (1997) Oppenheim conjecture. In: *Fields medalists lectures*, vol 5. World Sci Ser 20th Century Math. World Sci Publ, River Edge, pp 272–327
- Margulis GA (2000) Problems and conjectures in rigidity theory. In: *Mathematics: frontiers and perspectives*. AMS, Providence, pp 161–174
- Margulis GA, Qian N (2001) Local rigidity of weakly hyperbolic actions of higher rank real Lie groups and their lattices. *Ergodic Theory Dynam Syst* 21:121–164
- Margulis GA, Tomanov G (1994) Invariant measures for actions of unipotent groups over local fields of homogeneous spaces. *Invent Math* 116:347–392
- Mieczkowski D (2007) The first cohomology of parabolic actions for some higher-rank abelian groups and representation theory. *J Modern Dyn* 1:61–92
- Milnor J (1971) Introduction to algebraic K -theory. Princeton University Press, Princeton
- Mok N, Siu YT, Yeung SK (1993) Geometric superrigidity. *Invent Math* 113:57–83
- Mostow GD (1973) Strong rigidity of locally symmetric spaces. *Ann Math studies*, vol 78. Princeton University Press, Princeton
- Newhouse SE (1970) On codimension one Anosov diffeomorphisms. *Am J Math* 92:761–770
- Nițică V, Török A (1995) Cohomology of dynamical systems and rigidity of partially hyperbolic actions of higher rank lattices. *Duke Math J* 79:751–810
- Nițică V, Török A (1998) Regularity of the transfer map for cohomologous cocycles. *Ergodic Theory Dynam Syst* 18:1187–1209
- Nițică V, Török A (2001) Local rigidity of certain partially hyperbolic actions of product type. *Ergodic Theory Dynam Syst* 21:1213–1237
- Nițică V, Török A (2002) On the cohomology of Anosov actions. In: *Rigidity in dynamics and geometry*, Cambridge, 2000. Springer, Berlin, pp 345–361
- Nițică V, Török A (2003) Cocycles over abelian TNS actions. *Geom Dedicata* 102:65–90
- Oseledec VI (1968) A multiplicative ergodic theorem. Characteristic Lyapunov exponents of dynamical systems. *Trudy Mosk Mat Obsc* 19:179–210
- Parry W (1999) The Livšic periodic point theorem for non-abelian cocycles. *Ergodic Theory Dynam Syst* 19:687–701
- Pugh C, Shub M (1972) Ergodicity of Anosov actions. *Invent Math* 15:1–23
- Ratner M (1991a) On Ragunathan's measure conjecture. *Ann Math* 134:545–607
- Ratner M (1991b) Ragunathan's topological conjecture and distributions of unipotent flows. *Duke Math J* 63:235–280
- Ratner M (1995) Ragunathan's conjecture for Cartesians products of real and p -adic Lie groups. *Duke Math J* 77:275–382
- Rudolph D (1990) $\times 2$ and $\times 3$ invariant measures and entropy. *Ergodic Theory Dynam Syst* 10:395–406
- Schmidt K (1999) Remarks on Livšic' theory for nonabelian cocycles. *Ergodic Theory Dynam Syst* 19:703–721
- Selberg A (1960) On discontinuous groups in higher-dimensional symmetric spaces. In: *Contributions to*

- function theory. Inter colloq function theory, Bombay. Tata Institute of Fundamental Research, pp 147–164
- Smale S (1967) Differentiable dynamical systems. *Bull AMS* 73:747–817
- Spatzier R (1995) Harmonic analysis in rigidity theory. In: *Ergodic theory and its connections with harmonic analysis*. Alexandria, 1993. London Math Soc Lect Notes Ser, vol 205. Cambridge University Press, Cambridge, pp 153–205
- Veech WA (1986) Periodic points and invariant pseudo-measures for toral endomorphisms. *Ergodic Theory Dynam Syst* 6:449–473
- Verjovsky A (1974) Codimension one Anosov flows. *Bull Soc Math Mex* 19:49–77
- Weil A (1960) On discrete subgroups of Lie groups. I. *Ann Math* 72:369–384
- Weil A (1962) On discrete subgroups of Lie groups. II. *Ann Math* 75:578–602
- Weil A (1964) Remarks on the cohomology of groups. *Ann Math* 80:149–157
- Zimmer R (1984) *Ergodic theory and semi-simple groups*. Birkhäuser, Boston
- Zimmer R (1987) Actions of semi-simple groups and discrete subgroups. In: *Proc Inter Congress of Math* (1986). AMS, Providence, pp 1247–1258
- Zimmer R (1990) Infinitesimal rigidity of smooth actions of discrete subgroups of Lie groups. *J Differ Geom* 31: 301–322

Books and Reviews

- de la Harpe P, Valette A (1989) *La propriété (T) de Kazhdan pour les groupes localement compacts*. Astérisque 175
- Feres R (1998) *Dynamical systems and semi-simple groups: an introduction*. Cambridge tracts in mathematics, vol 126. Cambridge University Press, Cambridge
- Feres R, Katok A (2002) *Ergodic theory and dynamics of G-spaces*. In: *Handbook in dynamical systems*, vol 1A. Elsevier, Amsterdam, pp 665–763
- Gromov M (1988) *Rigid transformation groups*. In: Bernard D, Choquet-Bruhat Y (eds) *Géométrie Différentielle* (Paris, 1986). Hermann, Paris, pp 65–139. Travaux en Cours 33
- Kleinbock D, Shah N, Starkov A (2002) *Dynamics of subgroup actions on homogeneous spaces of Lie groups and applications to number theory*. In: *Handbook in dynamical systems*, vol 1A. Elsevier, Amsterdam, pp 813–930
- Knapp A (2002) *Lie groups beyond an introduction*, 2nd edn. Progress in mathematics, 140. Birkhäuser, Boston
- Raghunathan MS (1972) *Discrete subgroups of Lie groups*. Springer, Berlin
- Witte MD (2005) *Ratner's theorems on unipotent flows*. Chicago lectures in mathematics. University of Chicago Press, Chicago

Chaos and Ergodic Theory

Jérôme Buzzi

C.N.R.S. and Université Paris-Sud, Orsay, France

Article Outline

Glossary

Definition of the Subject

Introduction

Picking an Invariant Probability Measure

Tractable Chaotic Dynamics

Statistical Properties

Orbit Complexity

Stability

Untreated Topics

Future Directions

Bibliography

Glossary

For simplicity, definitions are given for a continuous or smooth self-map or diffeomorphism T of a compact manifold M .

Entropy, measure-theoretic (or: metric entropy) For an ergodic invariant probability measure μ , it is the smallest exponential growth rates of the number of orbit segments of given length, with respect to that length, after restriction to a set of positive measure. We denote it by $h(T, \mu)$. See ► “Entropy in Ergodic Theory” and section “Local Complexity” below.

Entropy, topological It is the exponential growth rates of the number of orbit segments of given length, with respect to that length. We denote it by $h_{\text{top}}(f)$. See ► “Entropy in Ergodic Theory” and section “Local Complexity” below.

Ergodicity A measure is ergodic with respect to a map T if given any measurable subset S which is invariant, i.e., such that $T^{-1}S = S$, either S or its complement has zero measure.

Hyperbolicity A measure is hyperbolic in the sense of Pesin if at almost every point no Lyapunov exponent is zero. See ► “Smooth Ergodic Theory”.

Kolmogorov typicality A property is *typical in the sense of Kolmogorov* for a topological space $\mathcal{F}$ of parametrized families $f = (f_t)_{t \in U}$, U being an open subset of $\mathbb{R}^d$ for some $d \geq 1$, if it holds for f_t for Lebesgue almost every t and topologically generic $f \in \mathcal{F}$.

Lyapunov exponents The Lyapunov exponents (► Smooth Ergodic Theory) are the limits, when they exist, $\lim_{n \rightarrow \infty} \frac{1}{n} \log \left\| (T^n)'(x) \cdot v \right\|$ where $x \in M$ and v is a non zero tangent vector to M at x . The Lyapunov exponents of an ergodic measure is the set of Lyapunov exponents obtained at almost every point with respect to that measure for all non-zero tangent vectors.

Markov shift (topological, countable state) It is the set of all infinite or bi-infinite paths on some countable directed graph endowed with the left-shift, which just translates these sequences.

Maximum entropy measure It is a measure μ which maximizes the measured entropy and, by the variational principle, realized the topological entropy.

Physical measure It is a measure μ whose basin, $\{x \in M : \forall \phi : M \rightarrow \mathbb{R} \text{ continuous } \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(f^k x) = \int \phi d\mu\}$ has nonzero volume.

Prevalence A property is *prevalent* in some complete metric, separable vector space X if it holds outside of a set N such that, for some Borel probability measure μ on X : $\mu(A + v) = 0$ for all $v \in N$. See (Christensen 1972; Hunt et al. 1992; Tsujii 1992).

Sensitivity on initial conditions The *sensitivity to initial conditions* on $X' \subset X$ if there exists a constant $\rho > 0$ such that for every $x \in X'$, there exists $y \in X$, arbitrarily close to x , and $n \geq 0$ such that $d(T^n y, T^n x) > \rho$.

Sinai–Ruelle–Bowen measures It is an invariant probability measure which is absolutely continuous along the unstable foliation (defined using the unstable manifolds of almost every $x \in M$, which are the sets $W^u(x)$, of points y such that $\lim_{n \rightarrow \infty} \frac{1}{n} \log d(T^{-n}y, T^{-n}x) < 0$).

Statistical stability T is statistically stable if the physical measures of nearby deterministic systems are arbitrarily close to the convex hull of the physical measures of T .

Stochastic stability T is stochastically stable if the invariant measures of the Markov chains obtained from T by adding a suitable, smooth noise with size $\epsilon \rightarrow 0$ are arbitrarily close to the convex hull of the physical measures of T .

Structural stability T is structurally stable if any S close enough to T is topologically the same as T : there exists a homeomorphism $h : M \rightarrow M$ such that $h \circ T = S \circ h$ (orbits are sent to orbits).

Subshift of finite type It is a closed subset Σ_F of $\Sigma = \mathcal{A}^{\mathbb{Z}}$ or $\Sigma = \mathcal{A}^{\mathbb{N}}$ where $\mathcal{A}$ is a finite set satisfying: $\Sigma_F = \{x \in \Sigma : \forall k < \ell : x_k x_{k+1} \dots x_\ell \notin F\}$ for some finite set F .

Topological genericity Let X be a Baire space, e.g., a complete metric space. A property is (topologically) generic in a space X (or holds for the (topologically) generic element of X) if it holds on a nonmeager set (or set of second Baire category), i.e., on a dense G_δ subset.

Definition of the Subject

Chaotic dynamical systems are those which present unpredictable and/or complex behaviors. The existence and importance of such systems has been known at least since Hadamard (1898) and Poincaré (1892), however it became well-known only in the sixties. We refer to (Bergelson 2006; Collet et al. 2005; Fiedler 2002; Guckenheimer 1979; Gutzwiller 1990; Hasselblatt and Katok 2002; Murray 2002; Puu 2000; Saari 2005; Starkov 2000) for the relevance of such dynamics in other fields, mathematical or not (see also ► “Ergodic Theory: Interactions with Combinatorics and Number Theory”, ► “Ergodic Theory: Fractal Geometry”). The numerical simulations of

chaotic dynamics can be difficult to interpret and to plan, even misleading, and the tools and ideas of mathematical dynamical system theory are indispensable. Arguably the most powerful set of such tools is ergodic theory which provides a statistical description of the dynamics by attaching relevant probability measures. In opposition to single orbits, the statistical properties of chaotic systems often have good stability properties. In many cases, this allows an understanding of the complexity of the dynamical system and even precise and quantitative statistical predictions of its behavior. In fact, chaotic behavior of single orbits often yields global stability properties.

Introduction

The word *chaos*, from the ancient Greek $\chi\alpha\omicron\sigma$, “shapeless void” (Hesiod 1987) and “raw confused mass” (Ovid 2005), has been used [*$\chi\alpha\omicron\sigma$ also inspired Van Helmont to create the word “gas” in the seventeenth century and this other thread leads to the molecular chaos of Boltzmann in the nineteenth century and therefore to ergodic theory itself.*] since a celebrated paper of Li and Yorke (1975) to describe evolutions which however deterministic and defined by rather simple rules, exhibit unpredictable or complex behavior.

Attempts at Definition

We note that, like many ideas (Stewart 1964), this is not captured by a single mathematical definition, despite several attempts (see, e.g., Blanchard et al. 2002; Glasner and Weiss 1993; Kolyada 2004; Ruelle 2003b) for some discussions as well as the monographs on chaotic dynamics (Arnol’d 1988; Arnol’d and Avez 1968; Brain and Berger 2001; Brin and Stuck 2002; Collet and Eckmann 2006; Eckmann and Ruelle 1985; Guckenheimer and Holmes 1990; Hasselblatt and Katok 2003; Robinson 2004; Viana 1997b; Young 1995). Let us give some of the most well-known definitions, which have been given mostly from the topological point of view, i.e., in the setting of a self-map $T : X \rightarrow X$ on a compact metric space whose distance is denoted by d :

T has *sensitivity to initial conditions* on $X' \subset X$ if there exists a constant $\rho > 0$ such that for every $x \in X'$, there exists $y \in X$, arbitrarily close to x with a finite separating time:

$$\exists n \geq 0 \quad \text{such that} \quad d(T^n y, T^n x) > \rho.$$

In other words, any uncertainty on the exact value of the initial condition x makes $T^n(x)$ completely unknown for n large enough. If X is a manifold, then *sensitivity to initial conditions in the sense of Guckenheimer (1979)* means that the previous phenomenon occurs for a set X' with nonzero volume.

T is *chaotic in the sense of Devaney (1989)* if it admits a dense orbit and if the periodic points are dense in X . It implies sensitivity to initial conditions on X .

T is *chaotic in the sense of Li and Yorke (1975)* if there exists an uncountable subset $X' \subset X$ of points, such that, for all $x \neq y \in X'$,

$$\begin{aligned} \liminf_{n \rightarrow \infty} d(T^n x, T^n y) &= 0 \quad \text{and} \\ \limsup_{n \rightarrow \infty} d(T^n x, T^n y) &> 0. \end{aligned}$$

T has *generic chaos* in the sense of Lasota (Piorek 1985) if the set

$$\left\{ (x, y) \in X \times X : \liminf_{n \rightarrow \infty} d(T^n x, T^n y) = 0 \right. \\ \left. < \limsup_{n \rightarrow \infty} d(T^n x, T^n y) \right\}$$

is topologically generic (see “[Glossary](#)”) in $X \times X$.

Topological chaos is also sometimes characterized by *nonzero topological entropy* (► [Entropy in Ergodic Theory](#)): there exist exponentially many orbit segments of a given length. This implies chaos in the sense of Li and Yorke by (Blanchard et al. 2002).

As we shall see ergodic theory describes a number of chaotic properties, many of them implying some or all of the above topological ones. The main such property for a smooth dynamical system, say a $C^{1+\alpha}$ -diffeomorphism of a compact manifold, is the existence of an invariant probability measure which is:

1. Ergodic (cannot be split) and aperiodic (not carried by a periodic orbit);
2. Hyperbolic (nearby orbits converge or diverge at a definite exponential rate);
3. Sinai–Ruelle–Bowen (as smooth as it is possible).

(For precise definitions we refer to ► [“Smooth Ergodic Theory”](#) or to the discussions below.) In particular such a situation implies nonzero entropy and sensitivity to initial condition of a set of nonzero Lebesgue measure (i.e., positive volume).

Before starting our survey in earnest, we shall describe an elementary and classical example, the full tent map, on which the basic phenomena can be analyzed in a very elementary way. Then, in section [“Picking an Invariant Probability Measure”](#), we shall give some motivations for introducing probability theory in the description of chaotic but deterministic systems, in particular the unpredictability of their individual orbits. We define two of the most relevant classes of invariant measures: the physical measures and those maximizing entropy. It is unknown in which generality these measures exist and can be analyzed but we describe in section [“Tractable Chaotic Dynamics”](#) the major classes of dynamics for which this has been done. In section [“Statistical Properties”](#) we describe some of the finer statistical properties that have been obtained for such good chaotic systems: sums of observables along orbits are statistically undistinguishable from sums of independent and identically distributed random variables. Section [“Orbit Complexity”](#) is devoted to the other side of chaos: the complexity of these dynamics and how, again, this complexity can be analyzed, and sometimes classified, using ergodic theory. Section [“Stability”](#) describes perhaps the most striking aspect of chaotic dynamics: the unstability of individual orbit is linked to various forms of stability of the global dynamics.

Finally we conclude by mentioning some of the most important topics that we could not address and we list some possible future directions.

Caveat. The subject-matter of this article is somewhat fuzzy and we have taken advantage of this to steer our path towards some of our favorite

theorems and to avoid the parts we know less (some of which are listed below). We make no pretense at exhaustivity neither in the topics nor in the selected results and we hope that our colleagues will excuse our shortcomings.

Remark 1 In this article we only consider *compact, smooth and finite-dimensional* dynamical systems *in discrete time*, i.e., defined by self-maps. In particular, we have omitted the natural and important variants applying to flows, e.g., evolutions defined by ordinary differential equations but we refer to the textbooks (see, e.g., Arnol'd 1988; Hasselblatt and Katok 2002; Katok and Hasselblatt 1995) for these.

Elementary Chaos: A Simple Example

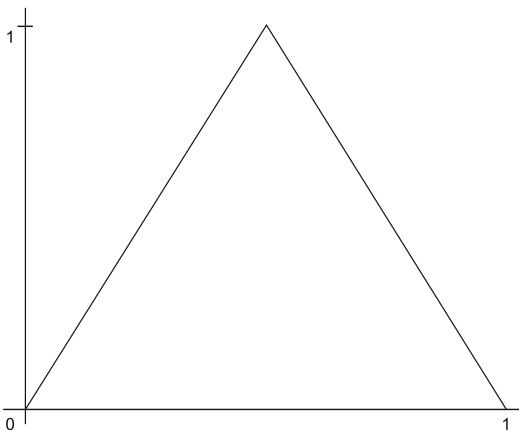
We start with a toy model: the full tent map T of Fig. 1. Observe that for any point $x \in [0, 1]$, $T^{-n}(x) = \{(\sigma(k, n)x + k) \cdot 2^{-n} : k = 0, 1, \dots, 2^n - 1\}$, where $\sigma(k, n) = \pm 1$. Hence $\cup_{n \geq 0} T^{-n}(x)$ is dense in $[0, 1]$. It easily follows that T exhibits *sensitive dependence to initial conditions*. Even worse in this example, the qualitative asymptotic behavior can be completely changed by this arbitrarily small perturbation: x may have a dense orbit whereas y is eventually mapped to a fixed point! This is Devaney chaos (Devaney 1989).

This kind of unstability was first discovered by J. Hadamard (1898) in his study of the geodesic flow (i.e., the frictionless movement of a point

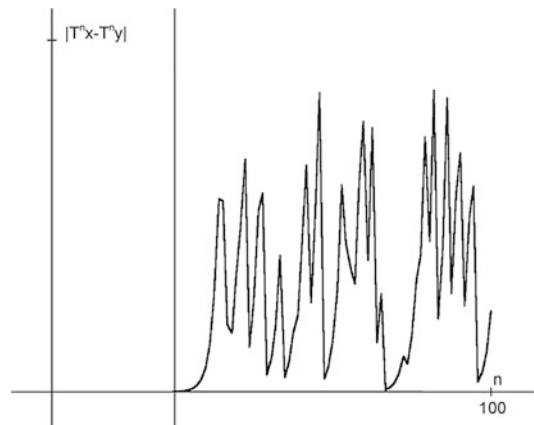
mass constrained to remain on a surface). At that time, such an unpredictability was considered a purely mathematical pathology, necessarily devoid of any physical meaning [Duhem qualified Hadamard's result as "an example of a mathematical deduction which can never be used by physics" (see pp. 206–211 in (Duhem 1906))!].

Returning to our tent map, we can be more quantitative. At any point $x \in [0, 1]$ whose orbit never visits $1/2$, the *Lyapunov exponent* $\lim_{n \rightarrow \infty} \frac{1}{n} \log |(T^n)'(x)|$ is $\log 2$. (See the "Glossary".) Such a positive Lyapunov exponent corresponds to infinitesimally close orbits getting separated exponentially fast. This can be observed in Fig. 2. Note how this exponential speed creates a rather sharp transition.

It follows in particular that experimental or numerical errors can grow very quickly to size 1, [For simple precision arithmetic the uncertainty is 10^{-16} which grows to size 1 in 38 iterations of T .] i.e., the approximate orbit may contain after a while *no information about the true orbit*. This casts a doubt on the reliability of simulations. Indeed, a simulation of T on most computers will suggest that all orbits quickly converge to 0, which is completely false [Such a collapse to 0 does really occur but only for a countable subset of initial conditions in $[0, 1]$ whereas the points with dense orbit make a subset of $[0, 1]$



Chaos and Ergodic Theory, Fig. 1 The graph of the full tent map $f(x) = 1 - |1 - 2x|$ over $[0, 1]$



Chaos and Ergodic Theory, Fig. 2 $|T^n(x) - T^n(y)|$ for $T(x) = 1 - |1 - 2x|$, $|x - y| = 10^{-12}$ and $0 \leq n \leq 100$. The vertical red line is at $n = 28$ and shows when $|T^n x - T^n y| \geq 0.5 \cdot 10^{-4}$ for the first time

with full Lebesgue measure (see below). This artefact comes from the way numbers are represented – and approximated – on the computer: multiplication by even integers tends to “simplify” binary representations. Thus the computations involved in drawing Fig. 2 cannot be performed too naively.]. Though somewhat atypical in its dramatic character, this failure illustrates the *unpredictability and instability* of individual orbits in chaotic systems.

Does this mean that all quantitative predictions about orbits of T are to be forfeited? Not at all, *if we are ready to change our point of view* and look beyond a single orbit. This can be seen easily in this case. Let us start with such a global analysis from the topological point of view. Associate to $x \in [0, 1]$, a sequence $i(x) = i = i_0 i_1 i_2 \dots$ of 0s and 1s according to $i_k = 0$ if $T^k x \leq 1/2$, $i_k = 1$ otherwise. One can check that [Up to a countable set of exceptions.] $\{i(x) : x \in [0, 1]\}$ is the set $\Sigma_2 := \{0, 1\}^{\mathbb{N}}$ of all infinite sequences of 0s and 1s and that at most one $x \in [0, 1]$ can realize a given sequence as $i(x)$.

Notice how the transformation f becomes trivial in this representation:

$$i(f(x)) = i_1 i_2 i_3 \dots \quad \text{if } i(x) = i_0 i_1 i_2 i_3 \dots$$

Thus f is represented by the simple and universal “left-shift” on sequences, which is denoted by σ . This representation of a rather general dynamical system by the left-shift on a space of sequences is called *symbolic dynamics* (► [Symbolic Dynamics](#)), (Lind and Marcus 1995).

This can be a very powerful tool. Observe for instance how here it makes obvious that we have complete *combinatorial freedom* over the orbits of T : one can easily build orbits with various asymptotic behaviors: if a sequence of Σ_2 contains all the finite sequences of 0s and 1s, then the corresponding point has a dense orbit; if the sequence is periodic, then the corresponding point is itself periodic, to give two examples of the *richness* of the dynamics.

More quantitatively, the number of distinct subsequences of length n appearing in sequences $i(x)$, $x \in [0, 1]$, is 2^n . It follows that the

topological entropy (► [Entropy in Ergodic Theory](#)) of T is $h_{\text{top}}(T) = \log 2$. [For the coincidence of the entropy and the Lyapunov exponent see below.] The positivity of the topological entropy can be considered as the signature of the complexity of the dynamics and considered as the definition, or at least the stamp, of a topologically chaotic dynamics.

Let us move on to a probabilistic point of view. Pick $x \in [0, 1]$ randomly according to, say, the uniform law in $[0, 1]$. It is then routine to check that $i(x)$ follows the $(1/2, 1/2)$ -Bernoulli law: the probability that, for any given k , $i_k(x) = 0$ is $1/2$ and the i_k s are independent. Thus $i(x)$, seen as a sequence of random 0 and 1 when x is subject to the uniform law on $[0, 1]$, is *statistically undistinguishable from coin tossing!* This important remark leads to quantitative predictions. For instance, the *strong law of large numbers* implies that, for Lebesgue-almost every $x \in [0, 1]$ (i.e., for all $x \in [0, 1]$ except in a set of Lebesgue measure zero), the fraction of the time spent in any dyadic interval $I = [k \cdot 2^{-N}, \ell \cdot 2^{-N}] \subset [0, 1]$, $k, \ell, N \in \mathbb{N}$, by the orbit of x ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \# \{0 \leq k < n : T^k x \in I\} \quad (1)$$

exists and is equal to the length, 2^{-N} , of that interval. [Eq. (1) in fact holds for any interval I . This implies that the orbit of almost every $x \in [0, 1]$ visits all subintervals of $[0, 1]$, i.e., the orbit is dense: in complete contradiction with the above mentioned numerical simulation!] More generally, we shall see that, if $\phi : [0, 1] \rightarrow \mathbb{R}$ is any continuous function, then, for Lebesgue almost every- x ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(T^k x) \quad \text{exists and is equal to} \quad \int \phi(x) dx. \quad (2)$$

Using strong *mixing properties* (► [Ergodicity and Mixing Properties](#)) of the Lebesgue measure under T , one can prove further properties, e.g., sensitivity on initial conditions in the sense of Guckenheimer [*The Lebesgue measure is weak-*

mixing: *Lebesgue-almost all couples of points $(x, y) \in [0, 1]^2$ get separated. Note that it is not true of every couple off the diagonal: Counterexamples can be found among couples (x, y) with $T^n x = T^n y$ arbitrarily close to 1.]* and study the fluctuations of the averages $\frac{1}{n} \sum_{k=0}^{n-1} \phi(T^k x)$ by the way of limit theorems.

The above analysis relied on the very special structure of T but, as we shall explain, the ergodic theory of differentiable dynamical systems shows that all of the above (and much more) holds in some form for rather general classes of chaotic systems. The different chaotic properties are independent in general (e.g., one may have topological chaos whereas the asymptotic behavior of almost all orbits is periodic) and the proofs can become much more difficult. We shall nonetheless be rewarded for our efforts by the discovery of unexpected links between chaos and stability, complexity and simplicity as we shall see.

Picking an Invariant Probability Measure

One could think that dynamical systems, such as those defined by self-maps of manifolds, being *completely deterministic*, have nothing to do with probability theory. There are in fact several motivations for introducing various invariant probability measures.

Statistical Descriptions

An abstract goal might be to enrich the structure: a smooth self-map is a particular case of a Borel self-map, hence one can canonically attach to this map its set of all invariant Borel probability measures [From now on all measures will be Borel probability measures except if it is explicitly stated otherwise.], or just the set of *ergodic* [A measure is ergodic if all measurable invariant subsets have measure 0 or 1. Note that arbitrary invariant measures are averages of ergodic ones, so many questions about invariant measures can be reduced to ergodic ones.] ones. By the Krylov–Bogoliubov theorem (see, e.g., Katok and Hasselblatt 1995), this set is non-empty for any continuous self-map of a compact space.

By the following fundamental theorem ▶ **Ergodic Theorems**, each such measure is the statistical description of some orbit:

Birkhoff Pointwise Ergodic Theorem

Let $(X, \mathcal{F}, \mu)$ be a space with a σ -field and a probability measure. Let $f: X \rightarrow X$ be a measure-preserving map, i.e., $f^{-1}(\mathcal{F}) \subset \mathcal{F}$ and $\mu \circ f^{-1} = \mu$. Assume ergodicity of μ (See the “**Glossary**”) and (absolute) integrability of $\phi: X \rightarrow \mathbb{R}$ with respect to μ . Then for μ -almost every $x \in X$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(f^k x) \quad \text{exists and is} \quad \int \phi \, d\mu.$$

This theorem can be interpreted as saying that “time averages” coincide almost surely with “ensemble averages” (or “phase space average”), i.e., that Boltzmann’s Ergodic Hypothesis of statistical mechanics (Gallavotti 1999) holds for dynamical systems that cannot be split in a measurable and non trivial way. [This indecomposability is however often difficult to establish. For instance, for the hard ball model of a gas it is known only under some generic assumption (see (Szász 2000) and the references therein).] We refer to (Krengel 1983) for background.

Remark 2 One should observe that the existence of the above limit is not at all obvious. In fact it often fails from other points of view. One can show that for the full tent map $T(x) = 1 - |1 - 2x|$ analyzed above and many functions ϕ , the set of points for which it fails is large both *from the topological point of view* (it contains a dense G_δ set) and *from the dimension point of view* (it has Hausdorff dimension 1 (Barreira and Schmeling 2000)). This is an important point: the introduction of invariant measures allows one to avoid some of the wilder pathologies.

To illustrate this let us consider the full tent map $T(x) = 1 - |1 - 2x|$ again and the two ergodic invariant measures: δ_0 (the Dirac measure concentrated at the fixed point 0) and the Lebesgue measure dx . In the first case, we obtain

a complex proof of the obvious fact that the time average at $x = 0$ (some set of full measure!) and the ensemble average with respect to δ_0 are both equal to $f(0)$. In the second case, we obtain a very general proof of the above Eq. (2).

Another type of example is provided by the contracting map $S : [0, 1] \rightarrow [0, 1]$, $S(x) = x/2$. S has a unique invariant probability measure, δ_0 . For Birkhoff theorem the situation is the same as that of T and δ_0 : it asserts only that the orbit of 0 is described by δ_0 .

One can understand Birkhoff theorem as a (first and rather weak) *stability result*: the time averages are independent of the initial condition, *almost surely with respect to μ* .

Physical Measures

In the above silly example S , much more is true than the conclusion of Birkhoff Theorem: *all* points of $[0, 1]$ are described by δ_0 . This leads to the definition of the *basin* of a probability measure μ for a self-map f of a space M :

$$\mathcal{B}(\mu) := \{x \in M : \forall \phi :$$

$$M \rightarrow \mathbb{R} \text{ continuous } \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(f^k x) = \int \phi d\mu \}.$$

If M is a manifold, then there is a notion of volume and one can make the following definition. A *physical measure* is a probability measure whose basin has nonzero volume in M . Say that a dynamical system $f : M \rightarrow M$ on a manifold has a *finite statistical description* if there exists finitely many invariant probability measures $\mu_1, \dots, \mu_n$ the union of whose basins is the whole of M , up to a set of zero Lebesgue measure.

Physical measures are among the main subject of interest as they are expected to be exactly those that are “experimentally visible”. Indeed, if $x_0 \in \mathcal{B}(\mu)$ and $\epsilon_0 > 0$ is small enough, then, by Lebesgue density theorem, a point x picked according to, say, the uniform law in the ball $B(x_0, \epsilon_0)$ of center x_0 and radius ϵ_0 , will be in $\mathcal{B}(\mu)$ with probability almost 1 and therefore its ergodic averages will be described by μ . Hence “experiments” can be expected to follow the physical measures and this is what is numerically

observed in most of the situations (see however the caveat in the discussion of the full tent map).

The existence of a finite statistical description (or even of a physical measure) is, as we shall see, not automatic nor routine to prove. Attracting periodic points as in the above silly example provide a first type of physical measures. Birkhoff ergodic theorem asserts that absolutely continuous ergodic invariant measures, usually obtained from some expansion property, give another class of physical measures. These contracting and expanding types can be combined in the class of *Sinai–Ruelle–Bowen measures* (Ledrappier 1984) which are the invariant measures absolutely continuous “along expanding directions” (see for the precise but technical definition ► [Smooth Ergodic Theory](#)). Any ergodic Sinai–Ruelle–Bowen measure which is ergodic and without zero Lyapunov exponent [*That is, the set of points $x \in M$ such that $\lim_{n \rightarrow \infty} \frac{1}{n} \log \|(f^n)'(x) \cdot v\| = 0$ for some $v \in T_x M$ has zero measure.*] is a physical measure. Conversely, “most” physical measures [*For counter-examples see* (Hofbauer and Keller 1990).] are of this type (Tsuji 2005; Vasquez 2007).

Measures of Maximum Entropy

For all parameters $t \in [3.96, 4]$, the quadratic maps $Q_t(x) = tx(1 - x)$, $Q_t : [0, 1] \rightarrow [0, 1]$, have nonzero topological entropy (de Melo and van Strien 1993) and exponentially many periodic points (Hofbauer 1985):

$$\lim_{n \rightarrow \infty} \frac{\#\{x \in [0, 1] : Q_t^n(x) = x\}}{e^{nh_{\text{top}}(Q_t)}} = 1.$$

On the other hand, by a deep theorem (Graczyk and Świątek 1997; Lyubich 1997) there is an open and dense subset of $t \in [0, 4]$, such that Q_t has a unique physical measure *concentrated on a periodic orbit*! Thus the physical measures can completely miss the topological complexity (and in particular the distribution of the periodic points). Hence one must look at other measures to get a statistical description of the complexity of such Q_t . Such a description is often provided by *measures of maximum entropy* μ_M whose

measured entropy [The usual phrases are “measure-theoretic entropy”, “metric entropy”.] (► [Entropy in Ergodic Theory](#)) satisfies:

$$h(f, \mu_M) = \sup_{\mu \in M(f)} h(f, \mu) = h_{\text{top}}(f).$$

$M(f)$ is the set of all invariant measures. [One can restrict this to the ergodic invariant measures without changing the value of the supremum (► [Entropy in Ergodic Theory](#)).] Equality 1 above is the *variational principle*: it holds for all continuous self-maps of compact metric spaces. One can say that the ergodic complexity (the complexity of f as seen by its *invariant measures*) captures the full topological complexity (defined by counting *all* orbits).

Remark 3 The variational principle implies the existence of “complicated invariant measures” as soon as the topological entropy is nonzero (see (Bonano and Collet 2006) for a setting in which this is of interest).

Maximum entropy measures do not always exist. However, if f is C^∞ smooth, then maximum entropy measures exist by a theorem of Newhouse (1989) and they indeed describe the topological complexity in the following sense. Consider the probability measures:

$$\mu_{n,\epsilon} := \frac{1}{n} \sum_{k=0}^{n-1} \sum_{x \in E(n,\epsilon)} \delta_{f^k x}$$

where $E(n, \epsilon)$ is an arbitrary (ϵ, n) -separated subset [See ► [Entropy in Ergodic Theory](#): $\forall x, y \in E(n, \epsilon) \ x \neq y \Rightarrow \exists 0 \leq k < n \ d(T^k x, T^k y) \geq \epsilon$.] of M with maximum cardinality. Then accumulation points for the weak star topology on the space of probability measures on M of $\mu_{n,\epsilon}$ when $n \rightarrow \infty$ and then $\epsilon \rightarrow 0$ are maximum entropy measures (Misiurewicz 1976a).

Let us quote two important additional properties, discovered by Margulis (2004), that often hold for the maximum entropy measures:

- The *equidistribution of periodic points* with respect to some maximum entropy measure μ_M :

$$\mu_M = \lim_{n \rightarrow \infty} \frac{1}{\#\{x \in X : x = f^n x\}} \sum_{x=f^n x} \delta_x.$$

- The *holonomy invariance* which can be loosely interpreted by saying that the past and the future are independent conditionally on the present.

Other Points of View

Many other invariant measures are of interest in various contexts and we have made no attempt at completeness: for instance, invariant measures maximizing dimension (Gatzouras and Peres 1997; Pesin 1997), or pressure in the sense of the thermodynamical formalism (Katok and Hasselblatt 1995; Ruelle 2004), or some energy (Anantharaman 2004; Contreras et al. 2001; Jenkinson 2007), or quasi-physical measures describing the dynamics around saddle-type invariant sets (Eckmann and Ruelle 1985) or in systems with holes (Chernov et al. 2000).

Tractable Chaotic Dynamics

The Palis Conjecture

There is, at this point, no general theory allowing the analysis of all dynamical systems or even of most of them despite many recent and exciting developments in the theory of generic C^1 -diffeomorphisms (Bonatti et al. 2005; Crovisier 2006a). In particular, the question of the generality in which physical measures exist remains open. One would like generic systems to have a *finite statistical description* (see section “[Physical Measures](#)”). This fails in some examples but these look exceptional and the following question is asked by Palis (2000):

Is it true that any dynamical system defined by a C^r -diffeomorphism on a compact manifold can be transformed by an arbitrarily small C^r -perturbation to another dynamical system having a finite statistical description?

This is completely open though widely believed [Observe, however, that such a statement is false for conservative diffeomorphisms with

high order smoothness as KAM theory implies stable existence of invariant tori foliating a subset of positive volume.]. Note that such a good description is not possible for all systems (see, e.g., Hofbauer and Keller 1990; Newhouse 1974). Note that one would really like to ask about unperturbed “typical” [The choice of the notion of typicality is a delicate issue. The Newhouse phenomenon shows that among C^2 -diffeomorphisms of multidimensional compact manifolds, one cannot use topological genericity and get a positive answer. Popular notions are prevalence and Kolmogorov genericity – see the “Glossary”.] dynamical systems in a suitable sense, but of course this is even harder.

One is therefore led to make simplifying assumptions: typically of small dimension, uniform expansion/contraction or geometry.

Uniformly Expanding/Hyperbolic Systems

The most easily analyzed systems are those with uniform expansion and/or contraction, namely the uniformly expanding maps and uniformly hyperbolic diffeomorphisms, see ► “Smooth Ergodic Theory”. [We require uniform hyperbolicity on the so-called chain recurrent set. This is equivalent to the usual Axiom-A and no-cycle condition.] An important class of example is obtained as follows. Consider $A : \mathbb{R}^d \rightarrow \mathbb{R}^d$, a linear map preserving $\mathbb{Z}^d$ (i.e., A is a matrix with integer coefficients in the canonical basis) so that it defines a map $\bar{A} : \mathbb{T}^d \rightarrow \mathbb{T}^d$ on the torus. If there is a constant $\Lambda > 1$ such that for all $v \in \mathbb{R}^d$, $\|A \cdot v\| \geq \Lambda \|v\|$ then $\bar{A}$ is a uniformly expanding map. If A has determinant ± 1 and no eigenvalue on the unit circle, then $\bar{A}$ is a uniformly hyperbolic diffeomorphism (► Smooth Ergodic Theory) (see also Brin and Stuck 2002; Katok and Hasselblatt 1995; Robinson 2004; Shub 1987). Moreover all C^1 -perturbations of the previous examples are again uniformly expanding or uniformly hyperbolic. [One can define uniform hyperbolicity for flows and an important class of examples is provided by the geodesic flow on compact manifolds with negative sectional curvature (Katok and Hasselblatt 1995).] These uniform systems are sometimes called “strongly chaotic”.

Remark 4 Mañé Stability Theorem (see below) shows that uniform hyperbolicity is a very natural notion. One can also understand on a more technical level uniform hyperbolicity as what is needed to apply an implicit function theorem in some functional space (see, e.g., Shub 1987).

The existence of a finite statistical description for such systems has been proved since the 1970s by Bowen, Ruelle and Sinai (Bowen and Ruelle 1975; Ruelle 1976; Ya 1972) (the expanding case is much simpler (Krzyżewski and Szlenk 1969)).

Theorem 1 Let $f : M \rightarrow M$ be a $C^{1+\alpha}$ map of a compact manifold. Assume f to be (i) a uniformly expanding map on M or (ii) a uniformly hyperbolic diffeomorphism.

- f admits a finite statistical description by ergodic and hyperbolic Sinai–Ruelle–Bowen measures (absolutely continuous in case (i)).
- f has finitely many ergodic maximum entropy measures, each of which makes f isomorphic to a finite state Markov chain. The periodic points are uniformly distributed according to some canonical average of these ergodic maximum entropy measures.
- f is topologically conjugate [Up to some negligible subset.] to a subshift of finite type (See the “Glossary”).

The construction of absolutely continuous invariant measures for a uniformly expanding map f can be done in a rather direct way by considering the pushed forward measures $\frac{1}{n} \sum_{k=0}^{n-1} f_*^k \text{Leb}$ and taking weak star limits while preventing the appearance of singularities, by, e.g., bounding some Hölder norm of the density using expansion and distortion of f .

The classical approach to the uniformly hyperbolic dynamics (Bowen 1975; Ruelle 2004; Shub 1987) is through symbolic dynamics and coding. Under the above hypothesis one can build a finite partition of M which is tailored to the dynamics (a Markov partition) so that the corresponding symbolic dynamics has a very simple structure: it is a full shift $\{1, \dots, d\}^{\mathbb{Z}}$, like in the example of

the full tent map, or a subshifts of finite type. The above problems can then be solved using the thermodynamical formalism inspired from the statistical mechanics of one-dimensional ferromagnets (Ruelle 1968): ergodic properties are obtained through the spectral properties of a suitable transfer operator acting on some space of regular functions, e.g., the Hölder-continuous functions defined over the symbolic dynamics with respect to the distance $d(x, y) := \sum_{n \in \mathbb{Z}} 2^{-n} 1_{x_n \neq y_n}$ where $1_{s \neq t}$ is 1 if $s \neq t$, 0 otherwise.

A recent development (Baladi and Tsujii 2007; Blank et al. 2002; Gouëzel and Liverani 2006) has been to find suitable Banach spaces to apply the transfer operator technique directly in the smooth setting, which not only avoids the complication of coding (or rather replace them with functional analytic preliminaries) but allows the use of the smoothness beyond Hölder-continuity which is important for finer ergodic properties.

Uniform expansion or hyperbolicity can easily be obstructed in a given system: a “bad” point (a critical point or a periodic point with multiplier with an eigenvalue on the unit circle) is enough. This leads to the study of other systems and has motivated many works devoted to relaxing the uniform hyperbolicity assumptions (Bonatti et al. 2005).

Pesin Theory

The most general such approach is *Pesin theory*. Let f be a $C^{1+\alpha}$ -diffeomorphism [It is an important open problem to determine to which extent Pesin theory could be generalized to the C^1 setting.] f with an ergodic invariant measure μ . By Oseledets Theorem ▶ [Smooth Ergodic Theory](#), for almost every x with respect to any invariant measure, the behavior of the differential $T_x f^n$ for n large is described by the Lyapunov exponents $\lambda_1, \dots, \lambda_d$ at x . Pesin is able to build charts around almost every orbit in which this asymptotic linear behavior describes that of f at the first iteration. That is, there are diffeomorphisms $\Phi_x : U_x \subset M \rightarrow V_x \subset \mathbb{R}^d$ with a “reasonable dependence on x ” such that the differential of $\Phi_{f^n x} \circ f^n \circ \Phi_x^{-1}$ at any point where it is defined is close to a diagonal matrix with entries $(e^{(\lambda_1 \pm \epsilon)n}, e^{(\lambda_2 \pm \epsilon)n}, \dots, e^{(\lambda_d \pm \epsilon)n})$.

In this full generality, one already obtains significant results:

- The entropy is bounded by the expansion:

$$h(f, \mu) \leq \sum_{i=1}^d \lambda_i^+(\mu) \text{ (Ruelle 1978)}$$

- At almost every point x , there are strong stable resp. unstable manifolds $W^{ss}(x)$, resp. $W^{uu}(x)$, coinciding with the sets of points y such that $d(T^n x, T^n y) \rightarrow 0$ exponentially fast when $n \rightarrow \infty$, resp. $n \rightarrow -\infty$. The corresponding holonomies are *absolutely continuous* (see, e.g., Brin 2001) like in the uniform case. This allows Ledrappier’s definition of Sinai–Ruelle–Bowen measures (Ledrappier 1984) in that setting.
- Equality in the above formula holds if and only if μ is a Sinai–Ruelle–Bowen measure (Ledrappier and Young 1985). More generally the entropy can be computed as $\sum_{i=1}^d \gamma_i(\mu) \lambda_i^+(\mu)$ where the γ_i are some fractal dimensions related to the exponents.

Under the only assumption of hyperbolicity (i.e., no zero Lyapunov exponent almost everywhere), one gets further properties:

- Existence of an hyperbolic measure which is not periodic forces $h_{\text{top}}(f) > 0$ [However $h(f, \mu)$ can be zero.] by (Katok 1980).
- μ is *exact dimensional* (Barreira et al. 1999; Young 1982): the limit $\lim_{r \rightarrow 0} \log \mu(B(x, r)) / \log r$ exists μ -almost everywhere and is equal to the Hausdorff dimension of μ (the infimum of the Hausdorff dimension of the sets with full μ -measure). This is deduced from a more technical “asymptotic product structure” property of any such measure.

For hyperbolic Sinai–Ruelle–Bowen measures μ , one can then prove, e.g.,:

- *local ergodicity* (Pesin 1976): μ has at most countably many ergodic components and μ -almost every point has a neighborhood whose Lebesgue-almost every point are contained in the basin of an ergodic component of μ .

- *Bernoulli* (Ornstein and Weiss 1988): Each ergodic component of μ is conjugate in a measure-preserving way, up to a period, to a Bernoulli shift, that is, a full shift $\{1, \dots, N\}^{\mathbb{Z}}$ equipped with a product measure. This in particular implies mixing and sensitivity on initial conditions for a set of positive Lebesgue measure.

However, establishing even such a weak form of hyperbolicity is rather difficult. The fragility of this condition can be illustrated by the result (Bochi 2002; Bochi and Viana 2005) that the topologically generic area-preserving surface C^1 -diffeomorphism is either uniformly hyperbolic or has Lebesgue almost everywhere vanishing Lyapunov exponents, hence is never non-uniformly hyperbolic! (but this is believed to be very specific to the very weak C^1 topology). Moreover, such weak hyperbolicity is not enough, with the current techniques, to build Sinai–Ruelle–Bowen measures or analyze maximum entropy measures only assuming non-zero Lyapunov exponents. Let us though quote two conjectures. The first one is from (Viana 1998) [We slightly strengthened Viana’s statement for expository reasons.]:

Conjecture 1 *Let f be a $C^{1+\epsilon}$ -diffeomorphism of a compact manifold. If Lebesgue-almost every point x has well-defined Lyapunov exponents in every direction and none of these exponents is zero, then there exists an absolutely continuous invariant σ -finite positive measure.*

The analogue of this conjecture has been proved for C^3 interval maps with unique critical point and negative Schwarzian derivative by Keller (1990), but only partial results are available for diffeomorphisms (Leplaideur 2004).

We turn to measures of maximum entropy. As we said, C^∞ smoothness is enough to ensure their existence but this is through a functional-analytic argument (allowed by Yomdin theory (Yomdin 1987)) which says nothing about their structure. Indeed, the following problem is open:

Conjecture 2 *Let f be a $C^{1+\epsilon}$ -diffeomorphism of a compact surface. If the topological entropy of f is nonzero then f has at most countably many ergodic invariant measures maximizing entropy.*

The analogue of this conjecture has been proved for $C^{1+\epsilon}$ interval maps (Buzzi 1997, 2000b, Buzzi). In the above setting a classical result of Katok shows the existence of uniformly hyperbolic compact invariant subsets with topological entropy arbitrarily close to that of f implying the existence of many periodic points:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \# \{x \in M : f^n(x) = x\} \geq h_{\text{top}}(f).$$

The previous conjecture would follow from the following one:

Conjecture 3 *Let f be a $C^{1+\epsilon}$ -diffeomorphism of a compact manifold. There exists an invariant subset $X \subset M$, carrying all ergodic measures with maximum entropy, such that the restriction $f|_X$ is conjugate to a countable state topological Markov shift (See the “Glossary”).*

Systems with Discontinuities

We now consider stronger assumptions to be able to build the relevant measures.

The simplest step beyond uniformity is to allow discontinuities, considering piecewise expanding maps. The discontinuities break the rigidity of the uniformly expanding situation. For instance, their symbolic dynamics are usually no longer subshifts of finite type though they still retain some “simplicity” in good cases (see Buzzi 2005).

To understand the problem in constructing the absolutely continuous invariant measures, it is instructive to consider the pushed forwards of a smooth measure. Expansion tends to keep the measure smooth whereas discontinuities may pile it up, creating non-absolute continuity in the limit. One thus has to check that expansion wins. In dimension 1, a simple fact resolves the argument: under a high enough iterate, one can make the expansion arbitrarily large everywhere, whereas a small interval can be chopped into at most two pieces.

Lasota and Yorke (1973) found a suitable framework. They considered C^2 interval maps with $|f'(x)| \geq \text{const} > 1$ except at finitely many points. They used the Ruelle transfer operator directly on the interval. Namely they studied

$$(L\phi)(x) = \sum_{y \in T^{-1}x} \frac{\phi(y)}{|T'(y)|}$$

acting on functions $\phi : [0, 1] \rightarrow \mathbb{R}$ with bounded variation and obtained the invariant density as the eigenfunction associated to the eigenvalue 1. One can then prove a *Lasota–Yorke inequality* (which might more accurately be called Doeblin–Fortet since it was introduced in the theory of Markov chains much earlier):

$$\|L\phi\|_{\text{BV}} \leq \alpha \|\phi\|_{\text{BV}} + \beta \|\phi\|_1 \quad (3)$$

where $\|\cdot\|_{\text{BV}}$, $\|\cdot\|_1$ are a strong and a weak norm, respectively and $\alpha < 1$ and $\beta < \infty$. One can then apply general theorems (Ionescu Tulcea and Marinescu 1950) or (Nussbaum 1970) (see (Baladi 2000a) for a detailed presentation of this approach and its variants). Here α can essentially be taken as 2 (reflecting the locally simple discontinuities) divided by the minimum expansion: so $\alpha < 1$, perhaps after replacing T with an iterate. In particular, the existence of a finite statistical description then follows (see Broise (1996a) for various generalizations and strengthenings of this result on the interval).

The situation in higher dimension is more complex for the reason explained above. One can obtain inequalities such as (3) on suitable if less simple functional spaces (see, e.g., Saussol 2000) but proving $\alpha < 1$ is another matter: discontinuities can get arbitrarily complex under iteration. (Buzzi 2001; Tsujii 2000b) show that indeed, in dimension 2 and higher, piecewise uniform expansion (with a finite number of pieces) is not enough to ensure a finite statistical description if the pieces of the map have only finite smoothness. In dimension 2, resp. 3 or more, piecewise real-analytic, resp. piecewise affine, is enough to exclude such examples (Buzzi 2000a; Tsujii 2000a), resp. (Tsujii 2001). (Cowieson 2002) has

shown that, for any $r > 1$, an open and dense subset of piecewise C^r and expanding maps have a finite statistical description.

Piecewise hyperbolic diffeomorphisms are more difficult to analyze though several results (conditioned on technical assumptions that can be checked in many cases) are available (Baladi and Gouëzel 2008; Chernov 1999; Sataev 1992; Young 1985).

Interval Maps with Critical Points

A more natural but also more difficult situation is a map for which the uniformity of the expansion fails because of the existence of critical points. [Note that, by a theorem of Mañé a circle map without critical points or indifferent periodic point is either conjugate to a rotation or uniformly expanding (Mañé 1985).]

A class which has been completely analyzed at the level of the above conjecture is that of *real-analytic families of maps of the interval* $f_t : [0, 1] \rightarrow [0, 1]$, $t \in I$, with a unique critical point, the main example being the quadratic family $Q_t(x) = tx(1-x)$ for $0 \leq t \leq 4$.

It is not very difficult to find quadratic maps with the following two types of behavior:

- (stable)** the orbit of Lebesgue-almost every $x \in [0, 1]$ tends to an attracting periodic orbit;
- (chaotic)** there is an absolutely continuous invariant probability measure μ whose basin contains Lebesgue-almost every $x \in [0, 1]$.

To realize the first it is enough to arrange the critical point to be periodic. One can easily prove that this stable behavior occurs on an open set of parameters –thus it is stable with respect to the parameter or the dynamical system. The second occurs for Q_4 with $\mu = dx/\sqrt{\pi x(1-x)}$. It is much more difficult to show that this chaotic behavior occurs for a set of parameters of positive Lebesgue measure. This is a theorem of Jakobson (1981) for the quadratic family (see for a recent variant (Young and Wang 2006)). Let us sketch two main ingredients of the various proofs of this theorem. The first is *inducing*: around Lebesgue-almost every point $x \in [0, 1]$ one tries to find a

time $\tau(x)$ and an interval $J(x)$ such that $f^{\tau(x)} : J(x) \rightarrow f^{\tau(x)}(J(x))$ is a map with good expansion and distortion properties. This powerful idea appears in many disguises in the non-uniform hyperbolic theory (see for instance (Hofbauer 1979; Young 1998)). The second ingredient is *parameter exclusion*: one removes the parameters at which a good inducing scheme cannot be built. More precisely one proceeds inductively, performing simultaneously the inducing and the exclusion, the good properties of the early stage of the inducing allowing one to control the measure of the parameters that need to be excluded to continue (Benedicks and Carleson 1985; Jakobson 1981). Indeed, the expansion established at a given stage allows to transfer estimates in the dynamical space to the parameter space.

Using methods from complex analysis and renormalization theory one can go much further and prove the following difficult theorems (actually the product of the work of many people, including Avila, Graczyk, Kozlovski, Lyubich, de Melo, Moreira, Shen, van Strien, Świątek), which in particular solves Palis conjecture in this setting:

Theorem 2 (Graczyk and Świątek 1997; Kozlovski et al. 2007; Lyubich 1997) *Stable maps (that is, such that Lebesgue almost every orbit converges to one of finitely many periodic orbits) form an open and dense set among C^r interval maps, for any $r \geq 2$. [In fact this is even true for polynomials].*

The picture has been completed in the *unimodal case* (that is, with a unique critical point):

Theorem 3 (Avila and Moreira 2005; Avila et al. 2003; Graczyk and Świątek 1997; Kozlovski 2003; Lyubich 1997) *Let $f_t : [0, 1] \rightarrow [0, 1]$, $t \in [t_0, t_1]$, be a real-analytic family of unimodal maps. Assume that it is not degenerate [f_{t_0} and f_{t_1} are not conjugate]. Then:*

- *The set of t such that f_t is chaotic in the above sense has positive Lebesgue measure;*

- *The set of t such that f_t is stable is open and dense;*
- *The remaining set of parameters has zero Lebesgue measure. [This set of “strange parameters” of zero Lebesgue measure has however positive Hausdorff dimension according to work of Avila and Moreira. In particular each of the following situations is realized on a set of parameters t of positive Hausdorff dimension: non-existence of the Birkhoff limit at Lebesgue-almost every point, the physical measure is δ_p for p a repelling fixed point, the physical measure is non-ergodic.]*

We note that the theory underlying the above theorem yields much more results, including a very paradoxical rigidity of typical analytic families as above. See (Avila and Moreira 2005).

Non-uniform Expansion/Contraction

Beyond the dimension 1, only partial results are available. The most general of those assume uniform contraction or expansion along some direction, restricting the non-uniform behavior to an invariant sub-bundle often one-dimensional, or “one-dimensional-like”.

A first, simpler situation is when there is a *dominated decomposition* with a uniformly expanding term: there is a continuous and invariant splitting of the tangent bundle, $T_\Lambda M = E^u \oplus E^s$, over some Λ an attracting set: for all unit vectors $v^u \in E^u$, $v^s \in E^s$,

$$\begin{aligned} \|(f^n)'(x) \cdot v^u\| &\geq C\lambda^n \quad \text{and} \\ \|(f^n)'(x) \cdot v^s\| &\leq C\mu^n \|(f^n)'(x) \cdot v^u\|. \end{aligned}$$

Standard techniques (pushing the Riemannian volume of a piece of unstable leaf and taking limits) allow the construction of Gibbs u -states as introduced by (Pesin and Sinai 1982).

Theorem 4 (Alves–Bonatti–Viana (2000)) *[A slightly different result is obtained in (Bonatti and Viana 2000).] Let $f : M \rightarrow M$ be a C^2 diffeomorphism with an invariant compact subset*

A. Assume that there is a dominated splitting $T_\Lambda M = E^u \oplus E^{cs}$ such that, for some $c > 0$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \prod_{k=0}^{n-1} \|f'(f^k) | E^{cs}\| \leq -c < 0$$

on a subset of Λ of positive Lebesgue measure. Then this subset is contained, up to a set of zero Lebesgue measure, in the union of the basins of finitely many ergodic and hyperbolic Sinai – Ruelle – Bowen measures.

The non-invertible, purely expansive version of the above theorem can be applied in particular to the following maps of the cylinder ($d \geq 16$, a is properly chosen close to 2 and α is small):

$$f(\theta, x) = (d\alpha \bmod 2\pi, a - x^2 + \epsilon \sin(\theta))$$

which are natural examples of maps with multi-dimensional expansion and critical lines considered by Viana (1997a). A series of works have shown that the above maps fit in the above non-uniformly expanding setting with a proper control of the critical set and hence can be thoroughly analyzed through variants of the above theorem (Alves 2000; Viana 1997a) and the references in (Alves 2006). For a, b properly chosen close to 2 and small ϵ , the following maps should be even more natural examples:

$$f(x, y) = (a - x^2 + \epsilon y, b - y^2 + \epsilon x). \quad (4)$$

However the inexistence of a dominated splitting has prevented the analysis of their physical measures. See (Buzzi 2000b; Buzzi) for their maximum entropy measures.

Cowieson and Young have used completely different techniques (thermodynamical formalism, Ledrappier–Young formula and Yomdin theory on entropy and smoothness) to prove the following result (see (Araujo and Tahzibi 2005) for related work):

Theorem 5 (Cowieson–Young 2005) *Let $f: M \rightarrow M$ be a C^∞ diffeomorphism of a compact manifold. Assume that f admits an attractor*

$A \subset M$ on which the tangent bundle has an invariant continuous decomposition $T_\Lambda M = E^+ \oplus E^-$ such that all vectors of $E^+ \setminus \{0\}$, resp. $E^- \setminus \{0\}$, have positive, resp. negative, Lyapunov exponents. Then any zero-noise limit measure μ of f is a Sinai–Ruelle–Bowen measure and therefore, if it is ergodic and hyperbolic, a physical measure.

One can hope, that typically, the latter ergodicity and hyperbolicity assumptions are satisfied (see, e.g., Baraviera and Bonatti 2003).

By pushing classical techniques and introducing new ideas for generic maps with one expanding and one weakly contracting direction, Tsujii has been able to prove the following generic result (which can be viewed as a 2-dimensional extension of some of the one-dimensional results above: one adds a uniformly expanding direction):

Theorem 6 (Tsujii 2005) *Let M be a compact surface. Consider the space of C^{20} self-maps $f: M \rightarrow M$ which admits directions that are uniformly expanded [More precisely, there exists a continuous, forward invariant cone field which is uniformly expanded under the differential of f .]*

Then existence of a finite statistical description is both topologically generic and prevalent in this space of maps.

Hénon-Like Maps and Rank One Attractors

In 1976, Hénon (1976) observed that the diffeomorphism of the plane

$$H_{a,b}(x, y) = (1 - ax^2 + y, bx)$$

seemed to present a “strange attractor” for $a = 1.4$ and $b = 0.3$, that is, the points of a numerically simulated orbit seemed to draw a set looking locally like the product of a segment with a Cantor set. This attractor seems to be supported by the unstable manifold $W^u(P)$ of the hyperbolic fixed point P with positive absciss. On the other hand, the Plykin classification (Plykin 2002) excluded the existence of a uniformly hyperbolic attractor for a dissipative surface diffeomorphism.

For almost twenty years the question of the existence of such an attractor (by opposition to an attracting periodic orbit with a very long period) remained open. Indeed, one knew since Newhouse that, for many such maps, there exist infinitely many such periodic orbits which are very difficult to distinguish numerically. But in 1991 Benedicks and Carleson succeeded in proposing an argument refining (with considerable difficulties) their earlier proof of Jakobson one-dimensional theorem and established the first part of the following theorem:

Theorem 7 (Benedicks–Carleson 1991) *For any $\epsilon > 0$, for $|b|$ small enough, there is a set A with $\text{Leb}(A) > 0$ satisfying: for all $a \in A$, there exists $z \in W^u(P)$ such that*

- The orbit of z is dense in $W^u(P)$;
- $\liminf_{n \rightarrow \infty} \frac{1}{n} \log \left\| (f^n)'(z) \right\| > 0$.

Further properties were then established, especially by Benedicks, Viana, Wang, Young (Benedicks and Viana 2001; Benedicks and Young 1993, 2000; Young and Wang 2001). Let us quote the following theorem of Wang and Young which includes the previous results:

Theorem 8 (Young and Wang 2001) *Let $T_{ab} : S^1 \times [-1, 1] \rightarrow S^1 \times [-1, 1]$ be such that*

- $T_{a0}(S^1 \times [-1, 1]) \subset S^1 \times \{0\}$;
- For $b > 0$, T_{ab} is a diffeomorphism on its image with

$$c^{-1}b \leq |\det T_{ab}(x, y)| \leq c \cdot b$$
for some $c > 1$ and all $(x, y) \in S^1 \times [-1, 1]$ and all (a, b) .

Let $f_a : S^1 \rightarrow S^1$ be the restriction of T_{a0} . Assume that $f = f_0$ satisfies:

- *Non-degenerate critical points: $f'(c) = 0 \Rightarrow f''(c) \neq 0$;*
- *Negative Schwarzian derivative: for all $x \in S^1$ non-critical, $f''(x)/f'(x)^3 - 3/2(f''(x)/f'(x))^2 < 0$;*
- *No indifferent or attracting periodic point, i.e., x such that $f^n(x) = x$ and $|(f^n)'(x)| \leq 1$;*

- *Misiurewicz condition: $d(f^n c, d) > c > 0$ for all $n \geq 1$ and all critical points c, d .*

Assume the following transversality condition on f at $a = 0$: for every critical point c , $(d/da)(f_a(c_a) - p_a) \neq 0$ if c_a is the critical point of f_a near c and p_a is the point having the same itinerary under f_a as $f(c)$ under c . Assume the following non-degeneracy of $T : f'_0(c) = 0 \Rightarrow \partial T_{00}(c, 0)/\partial y \neq 0$.

- *T_{ab} restricted to a neighborhood of $S^1 \times \{0\}$ has a finite statistical description by a number of hyperbolic Sinai–Ruelle–Bowen measures bounded by the number of critical points of f ;*
- *There is exponential decay of correlations and a Central Limit Theorem (see below) – except, in an obvious way, if there is a periodic interval with period > 1 ;*
- *There is a natural coding of the orbits that remains for ever close to $S^1 \times \{0\}$ by a closed invariant subset of a full shift.*

Very importantly, the above dynamical situation has been shown to occur near typical homoclinic tangencies: (Mora and Viana 1993) proved that there is an open and dense subset of the set of all C^3 families of diffeomorphisms unfolding a first homoclinic tangency such that the above holds. However (Palis and Yoccoz 2001) shows that the set of parameters with a Henon-like attractor has zero Lebesgue density at the bifurcation itself, at least under an assumption on the so-called stable and unstable Hausdorff dimensions. (Diaz et al. 1996) establishes positive density for another type of bifurcation. Furthermore (Moreira et al. 2001) has related the Hausdorff dimensions to the abundance of uniformly hyperbolic dynamics near the tangency.

Viana (1993) is able to treat situations with more than one contracting direction. More recently (Young and Wang 2008) has proposed a rather general framework, with easily checkable assumptions in order to establish the existence of such dynamics in various applications. See also (Guckenheimer et al. 2006; Wang and Young 2003) for applications.

Statistical Properties

The ergodic theorem asserts that time averages of integrable functions converge to phase space averages for any ergodic system. The speed of convergence is quite arbitrary in that generality (Krengel 1983) (only upcrossing inequalities seem to be available (Bishop 1967/1968; Hochman 2006)), however many results are available under very natural hypothesis as we are going to explain in this section. The underlying idea is that for sufficiently chaotic dynamics T and reasonably smooth observables ϕ , the time averages

$$A_n\phi(x) := \frac{1}{n} \sum_{k=0}^{n-1} \phi \circ T^k(x)$$

should behave as averages of independent and identically distributed random variables and therefore satisfy the classical limit theorems of probability theory.

The dynamical systems which are amenable to current technology are in a large part [*But other approaches are possible. Let us quote the work (Dolgopyat 2004a) on partially hyperbolic systems, for instance.*] those that can be reduced to the following type:

Definition 1 Let $T: X \rightarrow X$ be a nonsingular map on a probability, metric space $(X, \mathcal{B}, \mu, d)$ with bounded diameter, preserving the probability measure μ . This map is said to be **Gibbs–Markov** if there exists a countable (measurable) partition α of X such that:

1. For all $a \in \alpha$, T is injective on a and $T(a)$ is a union of elements of α .
2. There exists $\lambda > 1$ such that, for all $a \in \alpha$, for all points $x, y \in a$, $d(Tx, Ty) \geq \lambda d(x, y)$.
3. Let Jac be the *inverse* of the Jacobian of T . There exists $C > 0$ such that, for all $a \in \alpha$, for all points $x, y \in a$, $|1 - Jac(x)/Jac(y)| \leq Cd(Tx, Ty)$.

4. The map T has the “big image property”: $\inf_{a \in \alpha} \mu(Ta) > 0$.

Some piecewise expanding and C^2 maps are obviously Gibbs–Markov but the real point is that many dynamics can be reduced to that class by the use of inducing and tower constructions as in (Young 1998), in particular. This includes possibly piecewise uniformly hyperbolic diffeomorphisms, Collet–Eckmann maps of the interval (Baladi 2000a) (typical chaotic maps in the quadratic family), billiards with convex scatterers (Young 1998), the stadium billiard (Bálint and Gouëzel 2006), Hénon-like mappings (Young and Wang 2008).

We note that in many cases one is led to first analyze mixing properties through *decay of correlations*, i.e., to prove inequalities of the type (Baladi 2000a):

$$\left| \int_X \phi \cdot \psi \circ T^n d\mu - \int_X \phi d\mu \int_X \psi d\mu \right| \leq \|\phi\| \cdot \|\psi\|_w \cdot a_n \quad (5)$$

where $(a_n)_{n \geq 1}$ is some sequence converging to zero, e.g., $a_n = e^{-\lambda n}$, $1/n^\alpha$, $\dots$ and $\|\cdot\|$, $\|\cdot\|_w$ a strong and a weak norm (e.g., the variation norm and the L^1 norm). These rates of decay are often linked with return times statistics (Young 1999a). Rather general schemes have been developed to deduce various limit theorems such as those presented below from sufficiently quick decay of correlations (see notably (Liverani 1996) based on a dynamical variant of (Gordin 1969)).

Probabilistic Limit Theorems

The foremost limit property is the following:

Definition 2 A class $\mathcal{C}$ of functions $\phi: X \rightarrow \mathbb{R}$ is said to satisfy the Central Limit Theorem if the following holds: for all $\phi \in \mathcal{C}$, there is a number $\sigma = \sigma(\phi) > 0$ such that:

$$\lim_{n \rightarrow \infty} \mu \left(\left\{ x \in X : \frac{A_n \phi(x) - \int \phi d\mu}{\sigma n^{-1/2}} \leq t \right\} \right) = \int_{-\infty}^t e^{-x^2/2\sigma^2} \frac{dx}{\sqrt{2\pi}\sigma} \quad (6)$$

except for the degenerate case when $\phi(x) = \psi(Tx) - \phi(x) + \text{const.}$

The Central Limit Theorem can be seen in many cases as essentially a by-product of fast decay of correlations (Liverani 1996), i.e., if $\sum_{n \geq 0} a_n < \infty$ in the notations of Eq. (5). It has been established for Hölder-continuous observables for many systems together with their natural invariant measures including: uniformly hyperbolic attractors, piecewise expanding maps of the interval (Liverani 1995b), Collet–Eckmann unimodal maps on the interval (Keller and Nowicki 1992; Young 1992), piecewise hyperbolic maps (Chernov 1999), billiards with convex scatterers (Szász 2000), Hénon-like maps (Benedicks and Young 2000).

Remark 5 The classical Central Limit Theorem holds for square-integrable random variables (Nagaev 1957). For maps exhibiting *intermittency* (e.g., interval maps like $f(x) = x + x^{1+\alpha} \bmod 1$ with an indifferent fixed point at 0) the invariant density has singularities and the integrability condition is non longer automatic for smooth functions. One can then observe convergence to stable laws, instead of the normal law (Gouëzel 2004).

A natural question is the speed of the convergence in (6). The *Berry–Esseen inequality*:

$$\left| \mu \left(\left\{ \frac{A_n \phi(x) - \int \phi d\mu}{\sigma n^{-1/2}} \leq t \right\} \right) - \int_{-\infty}^t e^{-x^2/2\sigma^2} \frac{dx}{\sqrt{2\pi}\sigma} \right| \leq \frac{C}{n^{\delta/2}}$$

for some $\delta > 0$. It holds with $\delta = 1$ in the classical, probabilistic setting.

The *Local Limit Theorem* looks at a finer scale, asserting that for any finite interval $[a, b]$, any $t \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \sqrt{n} \mu \left(\{x \in X : nA_n \phi(x) \in [a, b] + \sqrt{n}t + n \int \phi d\mu\} \right) = |b - a| \frac{e^{-t^2/2\sigma^2}}{\sqrt{2\pi}\sigma}.$$

Both the Berry–Esseen inequality and the local limit theorem have been shown to hold for non-uniformly expanding maps (Gouëzel 2005) (also Broise 1996b; Rousseau-Egele 1983).

Almost Sure Results

It is very natural to try and describe the statistical properties of the averages $A_n(x)$ for almost every x , instead of the weaker above statements in probability over x . An important such property is the *almost sure invariance principle*. It asks for the discrete random walk defined by the increments of $\phi \circ T^n(x)$ to converge, after a suitable renormalization, to a Brownian motion. This has been proved for systems with various degrees of hyperbolicity (Denker and Philipp 1984; Dolgopyat 2004a; Hofbauer and Keller 1982; Melbourne and Nicol 2005). Another one is the almost sure Central Limit Theorem. In the independent case (e.g., if $X_1, X_2, \dots$ are independent and identically distributed random variables in L^2 with zero average and unit variance), the almost sure Central Limit Theorem states that, almost surely:

$$\frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} \delta_{\sum_{j=0}^{k-1} X_j / \sqrt{k}}$$

converges in law to the normal distribution. This implies, that, almost surely, for any $t \in \mathbb{R}$:

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} \delta \left\{ \sum_{j=0}^{k-1} X_j / \sqrt{k} \leq t \right\} \\ = \int_{-\infty}^t e^{-x^2/2\sigma^2} \frac{dx}{\sqrt{2\pi}\sigma}.$$

compare to (6).

A general approach is developed in (Chazottes and Gouëzel 2007), covering Gibbs–Markov maps and those that can be reduced to it. They show in particular that the dynamical properties needed for the classical Central Limit Theorem in fact suffice to prove the above almost invariance principle and even the almost sure version of the Central Limit Theorem (using general probabilistic results, see (Berkes and Csáki 2001; Yoshihara 2004)).

Other Statistical Properties

Essentially all the statistical properties of sums of independent identically distributed random variables can be established for tractable systems. Thus one can also prove large deviations (Kifer 1990; Liu et al. 2003; Young 1990), iterated law of the logarithm, etc. We note that the monograph (Collet and Eckmann 2006) contains a nice introduction to the current work in this area.

Orbit Complexity

The orbit complexity of a dynamical system $f: M \rightarrow M$ is measured by its topological and measured entropies. We refer to ► “Entropy in Ergodic Theory” for detailed definitions.

The Variational Principle

Bowen–Dinaburg and Katok formulae can be interpreted as meaning that the topological entropy counts the number of arbitrary orbits whereas the measured entropy counts the number of orbits relevant for the given measure. In most situations, and in particular for continuous self-map of compact metric spaces, the following *variational principle* holds:

$$h_{\text{top}}(f) = \sup_{\mu \in M(f)} h(f, \mu)$$

where $M(f)$ is the set of all invariant probability measures.

This is all the more striking in light of the fact that for many systems, the set of points which are typical from the point of view of ergodic theory [for instance, those x such that $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(T^k x)$ exists for all continuous functions ϕ .] is topologically negligible [A subset of a countable union of closed sets with empty interior; that is, meager or first Baire category.].

Strict Inequality

For a fixed invariant measure, one can only assert that $h(f, \mu) \leq h_{\text{top}}(f)$. One should be aware that this inequality may be strict even for a measure with full support. For instance, it is not difficult to check that the full tent map with $h_{\text{top}}(f) = \log 2$, admits ergodic invariant measures with full support and zero entropy.

There are also examples of dynamical systems preserving Lebesgue measure which have simultaneously positive topological entropy and zero entropy with respect to Lebesgue measure. That this occurs for C^1 surface diffeomorphisms preserving area is a simple consequence of a theorem of Bochi (2002) according to which, generically in the C^1 topology, such a diffeomorphism is either uniformly hyperbolic or with Lyapunov exponents Lebesgue-almost everywhere zero. [Indeed, it is easy to build such a diffeomorphism having both a uniformly hyperbolic compact invariant subset which will have robustly positive topological entropy and a non-degenerate elliptic fixed point which will prevent uniform hyperbolicity and therefore force all Lyapunov exponents to be zero. But Ruelle–Margulis inequality then implies that the entropy with respect to Lebesgue measure is zero.] Smooth examples also exist (Bolsinov and Taimanov 2000).

Remark 6 Algorithmic complexity (Li and Vitanyi 1997) suggests another way to look at orbit complexity. One obtains in fact in this way another formula for the entropy. However this

point of view becomes interesting in some settings, like extended systems defined by partial differential equations in unbounded space. Recently, (Bonano and Collet 2006) has used this approach to build interesting invariant measures.

Orbit Complexity on the Set of Measures

We have considered the entropies of each invariant measure separately, sharing only the common roof of topological entropy. One may ask how these different complexity sit together. A first answer is given by the following theorem in the symbolic and continuous setting. Let.

Theorem 9 (Downarowicz–Serafin 2003) *Let K be a Choquet simplex and $H : K \rightarrow \mathbb{R}$ be a convex function.*

Say that H is realized by a self-map $f : X \rightarrow X$ and its set $M(f)$ of f -invariant probability measures equipped with the weak star topology if the following holds. There exists an affine homeomorphism $\Psi : M(f) \rightarrow K$ such that, if $h : M(f) \rightarrow [0, \infty]$ is the entropy function, $H = h \circ \Psi$.

Then

- *H is realized by some continuous self-map of a compact space if and only if it is an increasing limit of upper semicontinuous and affine functions.*
- *H is realized by some subshift on a finite alphabet, i.e., by the left shift on a closed invariant subset Σ of $\{1, 2, \dots, N\}^{\mathbb{Z}}$ for some $N < \infty$, if and only if it is upper semi-continuous*

Thus, in both the symbolic and continuous settings it is possible to have a unique invariant measure with any prescribed entropy. This stands in contrast to surface $C^{1+\epsilon}$ -diffeomorphisms for which the set of the entropies of ergodic invariant measures is always the interval $[0, h_{\text{top}}(f)]$ as a consequence of (Katok 1980).

Local Complexity

Recall that the topological entropy can be computed as: $h_{\text{top}}(f) = \lim_{\epsilon \rightarrow 0} h_{\text{top}}(f, \epsilon)$ where:

$$h_{\text{top}}(f, \epsilon) := \lim_{n \rightarrow \infty} \frac{1}{n} \log s(\delta, n, X)$$

where $s(\delta, n, E)$ is the maximum cardinality of a subset S of E such that:

$$x \neq y \Rightarrow \exists 0 \leq k < n \quad d(f^k x, f^k y) \geq \epsilon$$

(see Bowen's formula of the topological entropy ► [Entropy in Ergodic Theory](#)). Likewise, the measure-theoretic entropy $h(T, \mu)$ of an ergodic invariant probability measure μ is $\lim_{\epsilon \rightarrow 0} h(f, \mu, \epsilon)$ where:

$$h(T, \epsilon) := \lim_{n \rightarrow \infty} \frac{1}{n} \log r(\delta, n, \mu)$$

where $r(\delta, n, \mu)$ is the minimum cardinality of $C \subset X$ such that

$$\mu(\{x \in X : \exists y \in C \text{ such that} \\ \forall 0 \leq k < n \quad d(f^k x, f^k y) < \epsilon\}) > 1/2.$$

One can ask *at which scales does entropy arise for a given dynamical system?*, i.e., how the above quantities $h(T, \epsilon)$, $h(T, \mu, \epsilon)$ converge when $\epsilon \rightarrow 0$.

An answer is provided by the *local entropy*. [This quantity was introduced by Misiurewicz (1976b) under the name conditional topological entropy and is called tail entropy by Downarowicz (2005).] For a continuous map f of a compact metric space X , it is defined as:

$$\begin{aligned} h_{\text{loc}}(f) &:= \lim_{\epsilon \rightarrow 0} h_{\text{loc}}(f, \epsilon) \quad \text{with} \\ h_{\text{loc}}(f, \epsilon) &:= \sup_{x \in X} h_{\text{loc}}(f, \epsilon, x) \quad \text{and} \\ h_{\text{loc}}(f, \epsilon, x) &:= \lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \sup \frac{1}{n} \log s(\delta, n, \{y \in X : \\ &\quad \forall k \geq 0 \quad d(f^k y, f^k x) < \epsilon\}) \end{aligned}$$

Clearly from the above formulas:

$$\begin{aligned} h_{\text{top}}(f) &\leq h_{\text{top}}(f, \delta) + h_{\text{loc}}(f, \delta) \quad \text{and} \\ h(f, \mu) &\leq h(f, \mu, \delta) + h_{\text{loc}}(f, \delta). \end{aligned}$$

Thus the local entropy bounds the defect in uniformity with respect to the measure of the pointwise limit $h(f, \mu) = \lim_{\delta \rightarrow 0} h(f, \mu, \delta)$. An exercise in topology shows that the local entropy therefore also bounds the defect in upper semi-continuity of $\mu \mapsto h(f, \mu)$. In fact, by a result of Downarowicz (2005) (extended by David Burguet to the non-invertible case), there is a *local variational principle*:

$$h_{\text{loc}}(f) = \lim_{\delta \rightarrow 0} \left(\sup_{\mu} h(f, \mu, \delta) - h(f, \mu) \right) \\ = \sup_{\mu} \limsup_{\nu \rightarrow \mu} h(f, \nu) - h(f, \mu)$$

for any continuous self-map f of a compact metric space.

The local entropy is easily bounded for smooth maps using Yomdin's theory:

Theorem 10 (Buzzi 1997) *For any C^r map f of a compact manifold, $h_{\text{loc}}(f) \leq \frac{d}{r} \log \sup_x \|f'(x)\|$. In particular, $h_{\text{loc}}(f) = 0$ if $r = \infty$.*

Thus C^∞ smoothness implies the existence of a maximum entropy measure (this was proved first by Newhouse) and the existence of *symbolic extension*: a subshift over a finite alphabet $\sigma : \Sigma \rightarrow \Sigma$ and a continuous and onto map $\pi : \Sigma \rightarrow M$ such that $\pi \circ \sigma = f \circ \pi$. More precisely,

Theorem 11 (Boyle, Fiebig, Fiebig (2002)) *Given a homeomorphism f of a compact metric space X , there exists a principal symbolic extension $\sigma : \Sigma \rightarrow \Sigma$, i.e., a symbolic extension such that, for every σ -invariant probability measure ν , $h(\sigma, \nu) = h(f, \nu \circ \pi^{-1})$, if and only if $h_{\text{loc}}(f) = 0$.*

We refer to (Boyle and Downarowicz 2004; Downarowicz and Newhouse 2005) for further results, including a realization theorem showing that the continuity properties of the measured entropy are responsible for the properties of symbolic extensions and also results in finite smoothness.

Global Simplicity

One can marvel at the power of mathematical analysis to analyze such complex evolutions. Of course another way to look at this is to remark that this analysis is possible *once this evolution has been fitted in a simple setting*: one had to move focus away from an individual, unpredictable orbit, of, say, the full tent map to the set of all the orbits of that map, which is essentially the set of all infinite sequences over two symbols: a very simple set indeed corresponding to full *combinatorial freedom* [(Weiss 2002) describes a weakening of this which holds for all positive entropy symbolic dynamics.]. The complete description of a given typical orbit requires an infinite amount of information, whereas the set of all orbits has a finite and very tractable definition. The complexity of the individual orbits is seen now as coming from purely random choices inside a simple structure.

The classical systems, namely uniformly expanding maps or hyperbolic diffeomorphisms of compact spaces, have a *simple* symbolic dynamics. It is not necessarily a full shift like for the tent map, but it is a subshift of finite type, i.e., a subshift obtained from a full shift by forbidding finitely many finite subwords. What happens outside of the uniform setting?

A fundamental example is provided by *piecewise monotone maps*, i.e., interval maps with finitely many critical points or discontinuities. The partition cut by these points defines a symbolic dynamics. This subshift is usually not of finite type. Indeed, the topological entropy taking arbitrary finite nonnegative values [For instance, the topological entropy of the β -transformation, $x \mapsto \beta x \bmod 1$, is $\log \beta$ for $\beta \geq 1$.], a representation that respects it has to use an uncountable class of models. In particular models defined by finite data, like the subshifts of finite type, cannot be generally adequate. However there are tractable “almost finite representations” in the following senses:

Most symbolic dynamics $\Sigma(T)$ of piecewise monotone maps T can be defined by finitely many infinite sequences, the *kneading invariants* of Milnor and Thurston: $\kappa_0^+, \kappa_1^-, \kappa_1^+, \dots$,

$\kappa_{d+1}^- \in \{0, \dots, d\}^{\mathbb{N}}$ if d is the number of critical/discontinuity points. [The kneading invariants are the suitably defined (left and right) itineraries of the critical/discontinuity points and endpoints.] Namely,

$$\Sigma(T) = \left\{ \alpha \in \{0, \dots, d\}^{\mathbb{N}} : \forall n \geq 0 \right. \\ \left. \kappa_{\alpha_n}^+ \leq \sigma^n \alpha \leq \kappa_{\alpha_{n+1}}^- \right\}$$

where $\leq$ is a total order on $\{0, \dots, d\}^{\mathbb{N}}$ making the coding $x \mapsto \alpha$ non-decreasing. Observe how the kneading invariants determine $\Sigma(T)$ in an effective way: knowing their first n symbols is enough to know the sequences of length n which begin sequences of $\Sigma(T)$. We refer to (de Melo and van Strien 1993) for the wealth of information that can be extracted from these kneading invariants following Milnor and Thurston (1988).

This form of global simplicity can be extended to other classes of non-uniformly expanding maps, including those like Eq. (4) using the notions of subshifts and puzzles of *quasi-finite type* (Buzzi 2005, 2006). This leads to the notion and analysis of *entropy-expanding maps*, a new open class of non-uniformly expanding maps admitting critical hypersurfaces, defined purely in terms of entropies including the otherwise untractable examples of Eq. (4).

A generalization of the representation of uniform systems by subshifts of finite type is provided by *strongly positive recurrent countable state Markov shifts*, a subclass of Markov shifts (See “Glossary”) which shares many properties with the subshifts of finite type (Boyle et al. 2006; Gurevic 1996; Gurevic and Savchenko 1998; Ruelle 2003a; Sarig 2001).

These “simple” systems admit a *classification result* which in particular identifies their measures with entropy close to the maximum (Boyle et al. 2006). Such a classification generalizes (Ladler and Marcus 1979). The “ideology” here is that complexity of individual orbits in a simple setting must come from randomness, but purely random systems are classified by their entropy according to Ornstein (1970).

Stability

By definition, chaotic dynamical systems have orbits which are unstable and numerically unpredictable. It is all the more surprising that, once one accepts to consider their dynamics globally, they exhibit very good stability properties.

Structural Stability

A simple form of stability is *structural stability*: a system $f: M \rightarrow M$ is structurally C^r -stable if any system g sufficiently C^r -close to f is topologically the same as f , formally: g is topologically conjugate, i.e., there is some homeomorphism [If h were C^1 , the conjugacy would imply, among other things, that for every p -periodic point: $\det((f^p)'(x)) = \det((g^p)'(h(p)))$, a much too strong requirement.] $h: M \rightarrow M$ mapping the orbits of f to those of g , i.e., $g \circ h = h \circ f$.

Andronov and Pontryaguin argued in the 1930’s that only such structurally stable systems are physically relevant. Their idea was that the model of a physical system is always known only to some degree of approximation, hence mathematical model whose structure depends on arbitrarily small changes should be irrelevant.

A first question is: *What are these structurally stable systems?* The answer is quite striking:

Theorem 12 (Mañé 1988) *Let $f: M \rightarrow M$ be a C^1 diffeomorphism of a compact manifold.*

f is structurally stable among C^1 -diffeomorphisms of M if and only if f is uniformly hyperbolic on its chain recurrent set. [A point x is chain recurrent if, for all $\epsilon > 0$, there exists a finite sequence $x_0, x_1, \dots, x_n$ such that $x_0 = x_n = x$ and $d(f(x_k), x_{k+1}) < \epsilon$. The chain recurrent set is the set of all chain recurrent points.]

A basic idea in the proof of the theorem is that failure of uniform hyperbolicity gives the opportunity to make an arbitrarily small perturbation contradicting the structural stability. In higher smoothness the required perturbation lemmas (e.g., the closing lemma (Arnaud 1998; Katok and Hasselblatt 1995; Mañé 1982)) are not available.

We note that uniform hyperbolicity without invertibility does not imply C^1 -stability (Przytycki 1977).

A second question is: *are these stable systems dense?* (So that one could offer structurally stable models for all physical situations). A deep discovery around 1970 is that this is not the case:

Theorem 13 (Abraham–Smale, Simon (Abraham and Smale 1970; Simon 1972))

For any $r \geq 1$ and any compact manifold M of dimension ≥ 3 , the set of uniformly hyperbolic diffeomorphisms is not dense in the space of C^r diffeomorphisms of M . [They use the phenomenon called “heterodimensional homoclinic intersections”].

Theorem 14 (Newhouse 1974) *For any $r \geq 2$, for any compact manifold M of dimension ≥ 2 , the set of uniformly hyperbolic diffeomorphisms is not dense in the space of C^r diffeomorphisms of M . More precisely, there exists a non-empty open subset in this space which contains a dense G_δ subset of diffeomorphisms with infinitely many periodic sinks. [So these diffeomorphisms have no finite statistical description.]*

Observe that it is possible that uniform hyperbolicity could be dense among surface C^1 -diffeomorphisms (this is the case for C^1 circle maps by a theorem of Jakobson (1981)).

In light of Mañé C^1 -stability theorem this implies that *structurally stable systems are not dense*, thus one can robustly see behaviors that are topologically modified by arbitrarily small perturbations (at least in the C^1 -topology)! So one needs to look beyond these and face that topological properties of relevant dynamical system are not determined from “finite data”. It is natural to ask whether the dynamics is almost determined by “sufficient data”.

Continuity Properties of the Topological Dynamics

Structural stability asks the topological dynamics to remain *unchanged* by a small perturbation. It is probably at least as interesting to ask it to *change continuously*. This raises the delicate question of

which topology should be put on the rather wild set of topological conjugacy classes. It is perhaps more natural to associate to the system a topological invariant taking value in a more manageable set and ask whether the resulting map is continuous.

A first possibility is Zeeman’s *Tolerance Stability Conjecture*. He associated to each diffeomorphism the set of all the closures of all of its orbits and he asked whether the resulting map is continuous on a dense G_δ subset of the class of C^r -diffeomorphisms for any $r \geq 0$. This conjecture remains open, we refer to (Crovisier 2006b) for a discussion and related progress.

A simpler possibility is to consider our favorite topological invariant, the topological entropy, and thus ask whether the dynamical complexity as measured by the entropy is a stable phenomenon. $f \mapsto h_{\text{top}}(f)$ is lower semicontinuous for f among C^0 maps of the interval [On the set of interval maps with a bounded number of critical points, the entropy is continuous (Misiurewicz 1995). Also $t \mapsto h_{\text{top}}(Q_t)$ is non-decreasing by complex arguments (de Melo and van Strien 1993), though it is a non-smooth function.] (Misiurewicz 1979) and for f among $C^{1+\epsilon}$ -diffeomorphisms of a compact surface (Katok 1980). [It is an important open question whether this actually holds for C^1 -diffeomorphisms. It fails for homeomorphisms (Rees 1981).] In both cases, one shows the existence of structurally stable invariant uniformly expanding or hyperbolic subsets with topological entropy close to that of the whole dynamics. On the other hand $f \mapsto h_{\text{top}}(f)$ is upper semicontinuous for C^∞ maps (Newhouse 1989; Yomdin 1987).

Statistical Stability

Statistical stability is the property that *deterministic perturbations* of the dynamical system cause only small changes in the physical measures, usually with respect to the weak star topology on the space of measures. When the physical measure μ_g is uniquely defined for all systems g near f , statistical stability is the continuity of the map $g \mapsto \mu_g$ thus defined.

Statistical stability is known in the uniform setting and also in the piecewise uniform case,

provided the expansion is strong enough (Baladi and Young 1993; Blank and Keller 1997; Keller and Liverani 1999) (otherwise there are counterexamples, even in the one-dimensional case).

It also holds in some non-uniform settings without critical behavior, in particular for maps with dominated splitting satisfying robustly a separation condition between the positive and negative Lyapunov exponents (Vasquez 2007) (see also (Alves and Viana 2002)).

However statistical instability occurs in dynamics with critical behaviors (Avila, 2007, personal communication):

Theorem 15 *Consider the quadratic family $Q_t(x) = tx(1 - x)$, $t \in [0, 4]$. Let $I \subset [0, 4]$ be the full measure subset of $[0, 4]$ of good parameters t such that in particular Q_t admits a physical measure (necessarily unique). For $t \in I$, let μ_t be this physical measure.*

Lebesgue almost every $t \in I$ such that μ_t is not carried by a periodic orbit [At such parameters, statistical stability is easily proved.], is a discontinuity point of the map $t \mapsto \mu_t$ [$M([0, 1])$ is equipped with the vague topology.]. However, for Lebesgue almost every t , there is a subset $I_t \subset I$ for which t is a Lebesgue density point [t is a Lebesgue density point if for all $r > 0$, the Lebesgue measure $m(I_t \cap [t - r, t + r]) > 0$ and $\lim_{\epsilon \rightarrow 0} m(I_t \cap [t - \epsilon, t + \epsilon]) / 2\epsilon = 1$.] and such that $\mu : I_t \rightarrow M([0, 1])$ is continuous at t .

Stochastic Stability

A physically motivated and a technically easier approach is to study stability of the physical measure under *stochastic perturbations*. For simplicity let us consider a diffeomorphism of a compact subset of $\mathbb{R}^d$ allowing for a direct definition of *additive noise*. Let $\psi(x)dx$ be an absolutely continuous probability law with compact support. [Sometimes additional properties are required of the density, e.g., $\psi(x) = \phi(x)1_B$ where 1_B is the characteristic function of the unit ball and $C^{-1} \leq \phi(x) \leq C$ for some $C < \infty$.] For $\epsilon > 0$, consider the Markov chain f_ϵ with state space M and transition probabilities:

$$P_\epsilon(x, A) = \int_A \psi\left(\frac{y - f(x)}{\epsilon}\right) \frac{dy}{\epsilon^d}.$$

The evolution of measures is given by: $(f_\epsilon \mu)(A) = \int_M P_\epsilon(x, A) d\mu$. Under rather weak irreducibility assumptions on f , f_ϵ has a *unique invariant measure* μ_ϵ (contrarily to f) and μ_ϵ is absolutely continuous. When f has a unique physical measure μ , it is said to be *stochastically stable* if $\lim_{\epsilon \rightarrow 0} \mu_\epsilon = \mu$ in the appropriate topology (the weak star topology unless otherwise specified).

It turns out that stochastic stability is a rather common property of Sinai–Ruelle–Bowen measures. It holds not only for uniformly expanding maps or hyperbolic diffeomorphisms (Kifer 1977; Young 1986), but also for most interval maps (Baladi and Young 1993), for partially hyperbolic systems of the type $E^u \oplus E^{cs}$ or Hénon-like diffeomorphisms (Alves and Araujo 2003; Araujo and Tahzibi 2005; Benedicks and Viana 2006; Blank 1989). We refer to the monographs (Baladi 2000a; Blank 1997) for more background and results.

Untreated Topics

For reasons of space and time, many important topics have been left out of this article. Let us list some of them.

Other phenomena related to chaotic dynamics have been studied: entrance times (Collet 1996; Collet and Galves 1995), spectral properties and dynamical zeta functions (Baladi 2000a), escape rates (Eckmann and Ruelle 1985), dimension (Pesin 1997), differentiability of physical measures with respect to parameters (Dolgopyat 2004b; Ruelle 2005), entropies and volume or homological growth rates (Gromov 1987; Hua et al. 2006; Yomdin 1987).

As far as the structure of the setting is concerned, one can go beyond maps or diffeomorphisms or flows and study: more general group actions (Hasselblatt and Katok 2002); holomorphic and meromorphic structures (Carleson and Gamelin 1993) and the references therein; symplectic or volume-preserving (de la Llave 2001; Hofer and Zehnder 1994) and in

particular the Pugh–Shub program around stable ergodicity of partially hyperbolic systems (Pugh and Shub 1999); random iterations (Arnold 1998; Kifer 1986, 1988).

A number of important problems have motivated the study of special forms of chaotic dynamics: equidistribution in number theory (Eskin and McMullen 1993; Furstenberg 1981; Host and Kra 2005) and geometry (Starkov 2000); quantum chaos (Anantharaman and Nonnenmacher 2007; Gutzwiller 1990); chaotic control (Saperstone and Yorke 1971); analysis of algorithms (Vallée 2006).

We have also omitted the important problem of *applying the above results*. Perhaps because of the lack of a general theory, this can often be a challenge (see for instance (Tucker 1999) for the already complex problem of verifying *uniform hyperbolicity* for a singular flow). Liverani has shown how theoretical results can lead to precise and efficient estimates for the toy model of piecewise expanding interval maps (Liverani 2001). Ergodic theory implies that, in some settings at least, *adding noise may make some estimates more precise* (see Kifer 1997). We refer to (Fiedler 2001) and the reference therein.

Future Directions

We conclude this article by a (very partial) selection of open problems.

General Theory

In dimension 1, we have seen that the analogue of the Palis conjecture (see above) is established (Theorem 2). However the description of the typical dynamics in Kolmogorov sense is only known in the unimodal, non-degenerate case by Theorem 3. Indeed, results like (Bruin et al. 1996) suggest that the multimodal picture could be more complex.

In higher dimensions, our understanding is much more limited. As far as a general theory is concerned, a deep problem is the paucity of results on the generic dynamics in C^r smoothness with $r > 1$. The remarkable current progress in generic dynamics (culminating in the proof of the weak Palis conjecture, see (Crovisier 2006a), the references therein and (Bonatti et al. 2005) for

background) seems restricted to the C^1 topology because of the lack of fundamental tools (e.g., closing lemmas) in higher smoothness. But Pesin theory requires higher smoothness at least technically. This is not only a hard technical issue but generic properties of physical measures, when they have been analyzed are often completely different between the C^1 case and higher smoothness (Bochi and Viana 2005).

Physical Measures

In higher dimensions, Benedicks and Carleson analysis of the Hénon map has given rise to a rather general theory of Hénon-like maps and more generally of the dynamical phenomena associated to homoclinic tangencies. However, the proofs are extremely technical. Could they be simplified? Current attempts like (Young and Wang 2008) center on the introduction of a simpler notion of critical points, possibly a non-inductive one (Pujals and Rodríguez-Hertz 2007).

Can this Hénon theory be extended to the weakly dissipative situation? to the conservative situation (for which the *standard map* is a well-known example defying analysis)? In the strongly dissipative setting, what are the typical phenomena on the complement of the Benedicks–Carleson set of parameters?

From a global perspective one of the main questions is the following:

Can infinitely many sinks coexist for a large set of parameters in a typical family or is Newhouse phenomenon atypical in the Kolmogorov or prevalent sense?

This seems rather unlikely (see however (Araujo 2001)).

Away from such “critical dynamics”, there are many results about systems with dominated splitting satisfying additional conditions. Can these conditions be weakened so they would be satisfied by typical systems satisfying some natural conditions (like robust transitivity)? For instance:

Could one analyze the physical measures of volume-hyperbolic systems?

A more specific question is whether Tsujii’s striking analysis of surface maps with one uniformly expanding direction can be extended to higher dimensions? can one weaken the

uniformity of the expansion? The same questions for the corresponding invertible situation is considered in (Cowieson and Young 2005).

Maximum Entropy Measures and Topological Complexity

As we explained, C^∞ smoothness, by a Newhouse theorem, ensures the existence of maximum entropy measures, making the situation a little simpler than with respect to physical measures. This existence results allow in particular an easy formulation of the problem of the typicality of hyperbolicity:

Are maximum entropy ergodic measures of systems with positive entropy hyperbolic for most systems?

A more difficult problem is that of the finite multiplicity of the maximum entropy measures. For instance:

Do typical systems possess finitely many maximum entropy ergodic measures?

More specifically, can one prove intrinsic ergodicity (i.e. uniqueness of the measure of maximum entropy) for an isolated homoclinic class of some diffeomorphisms (perhaps C^1 -generic)? Can a generic C^1 -diffeomorphism carry an infinite number of homoclinic classes, each with topological entropy bounded away from zero?

A perhaps more tractable question, given the recent progress in this area: Is a C^1 -generic partially hyperbolic diffeomorphisms, perhaps with central dimension 1 or 2, intrinsically ergodic?

We have seen how uniform systems have simple symbolic dynamics, i.e., subshifts of finite type, and how interval maps and more generally entropy-expanding maps keep some of this simplicity, defining subshifts or puzzles of quasi-finite type (Buzzi 2005; Buzzi 2006). (Young and Wang 2001) have defined symbolic dynamics for topological Hénon-like map which seems close to that of that of a one-dimensional system.

Can one describe Wang and Young symbolic dynamics of Hénon-like attractors and fit it in a class in which uniqueness of the maximum entropy measure could be proved?

More generally, can one define nice combinatorial descriptions, for surface diffeomorphisms? Can one formulate variants of the entropy-

expansion condition [*For instance building on our “entropy-hyperbolicity”*] of (Buzzi 2000b, Buzzi), that would be satisfied by a large subset of the diffeomorphisms?

Another possible approach is illustrated by the pruning front conjecture of (Cvitanović et al. 1988) (see also (de Carvalho and Hall 2002; Ishii 1997)). It is an attempt to build a combinatorial description by trying to generalize the way that, for interval maps, kneading invariants determine the symbolic dynamics by considering the bifurcations from a trivial dynamics to an arbitrary one.

We hope that our reader has shared in our fascination with this subject, the many surprising and even paradoxical discoveries that have been made and the exciting current progress, despite the very real difficulties both in the analysis of such non-uniform systems as the Henon map and in the attempts to establish a general (and practical) ergodic theory of chaotic dynamical systems.

Acknowledgments I am grateful for the advice and/or comments of the following colleagues: P. Collet, J.-R. Chazottes, and especially S. Ruelle. I am also indebted to the anonymous referee.

Bibliography

Primary Literature

- Aaronson J, Denker M (2001) Local limit theorems for partial sums of stationary sequences generated by Gibbs–Markov maps. *Stochastics Dyn* 1:193–237
- Abdenur F, Bonatti C, Crovisier S (2006) Global dominated splittings and the C^1 newhouse phenomenon. *Proc Am Math Soc* 134(8):2229–2237
- Abraham R, Smale S (1970) Nongenericity of Ω -stability. In: *Global analysis*, vol XIV. *Proc Sympos Pure Math*. American Mathematical Society, Providence
- Alves JF (2000) SRB measures for non-hyperbolic systems with multidimensional expansion. *Ann Sci Ecole Norm Sup* 33(4):1–32
- Alves JF (2006) A survey of recent results on some statistical features of non-uniformly expanding maps. *Discrete Contin Dyn Syst* 15:1–20
- Alves JF, Araujo V (2003) Random perturbations of non-uniformly expanding maps. *Geometric methods in dynamics*. I. *Astérisque* 286:25–62
- Alves JF, Viana M (2002) Statistical stability for robust classes of maps with non-uniform expansion. *Ergodic Theory Dynam Syst* 22:1–32

- Alves JF, Bonatti C, Viana M (2000) SRB measures for partially hyperbolic systems whose central direction is mostly expanding. *Invent Math* 140:351–398
- Anantharaman N (2004) On the zero-temperature or vanishing viscosity limit for certain Markov processes arising from Lagrangian dynamics. *J Eur Math Soc (JEMS)* 6:207–276
- Anantharaman N, Nonnenmacher S (2007) Half delocalization of the eigenfunctions for the Laplacian on an Anosov manifold. *Ann Inst Fourier* 57:2465–2523
- Araujo V (2001) Infinitely many stochastically stable attractors. *Nonlinearity* 14:583–596
- Araujo V, Pacifico MJ (2006) Large deviations for non-uniformly expanding maps. *J Stat Phys* 125:415–457
- Araujo V, Tahzibi A (2005) Stochastic stability at the boundary of expanding maps. *Nonlinearity* 18: 939–958
- Arnaud MC (1998) Le “closing lemma” en topologie C^1 . *Mem Soc Math Fr (NS)* 74
- Arnol'd VI (1988) Geometrical methods in the theory of ordinary differential equations. *Grundlehren der Mathematischen Wissenschaften*, vol 250, 2nd edn. Springer, New York
- Arnol'd VI, Avez A (1968) Ergodic problems of classical mechanics. W.A. Benjamin, New York
- Arnold L (1998) Random dynamical systems. Springer monographs in mathematics. Springer, Berlin
- Avila A, Moreira CG (2005) Statistical properties of unimodal maps: the quadratic family. *Ann Math* 161(2):831–881
- Avila A, Lyubich M, de Melo W (2003) Regular or stochastic dynamics in real analytic families of unimodal maps. *Invent Math* 154:451–550
- Baladi V (2000a) Positive transfer operators and decay of correlations. In: *Advanced series in nonlinear dynamics*, vol 16. World Scientific Publishing, River Edge
- Baladi V, Gouëzel S (2008) Good Banach spaces for piecewise hyperbolic maps via interpolation. preprint arXiv:0711.1960. Available from <http://www.arxiv.org>
- Baladi V, Ruelle D (1996) Sharp determinants. *Invent Math* 123:553–574
- Baladi V, Tsujii M (2007) Anisotropic Hölder and Sobolev spaces for hyperbolic diffeomorphisms. *Ann Inst Fourier (Grenoble)* 57:127–154
- Baladi V, Young LS (1993) On the spectra of randomly perturbed expanding maps. *Commun Math Phys* 156: 355–385. Erratum: *Comm Math Phys* 166(1): 219–220, 1994
- Baladi V, Kondah A, Schmitt B (1996) Random correlations for small perturbations of expanding maps, English summary. *Random Comput Dynam* 4:179–204
- Bálint P, Gouëzel S (2006) Limit theorems in the stadium billiard. *Commun Math Phys* 263:461–512
- Baraviera AT, Bonatti C (2003) Removing zero Lyapunov exponents, English summary. *Ergodic Theory Dynam Syst* 23:1655–1670
- Barreira L, Schmeling J (2000) Sets of “non-typical” points have full topological entropy and full Hausdorff dimension. *Israel J Math* 116:29–70
- Barreira L, Pesin Y, Schmeling J (1999) Dimension and product structure of hyperbolic measures. *Ann Math* 149(2):755–783
- Benedicks M, Carleson L (1985) On iterations of $1-ax^2$ on $(-1,1)$. *Ann Math* (2) 122(1):1–25
- Benedicks M, Carleson L (1991) The dynamics of the Hénon map. *Ann Math* 133(2):73–169
- Benedicks M, Viana M (2001) Solution of the basin problem for Hénon-like attractors. *Invent Math* 143:375–434
- Benedicks M, Viana M (2006) Random perturbations and statistical properties of Hénon-like maps. *Ann Inst H Poincaré Anal Non Linéaire* 23:713–752
- Benedicks M, Young LS (1993) Sinai–Bowen–Ruelle measures for certain Hénon maps. *Invent Math* 112: 541–576
- Benedicks M, Young LS (2000) Markov extensions and decay of correlations for certain Hénon maps. *Géométrie complexe et systèmes dynamiques*, Orsay, 1995. *Astérisque* 261:13–56
- Bergelson V (2006) Ergodic Ramsey theory: a dynamical approach to static theorems. In: *International congress of mathematicians*, vol II. European Mathematical Society, Zürich, pp 1655–1678
- Berkes I, Csáki E (2001) A universal result in almost sure central limit theory. *Stochastic Process Appl* 94(1): 105–134
- Bishop E (1967/1968) A constructive ergodic theorem. *J Math Mech* 17:631–639
- Blanchard F, Glasner E, Kolyada S, Maass A (2002) On Li-Yorke pairs, English summary. *J Reine Angew Math* 547:51–68
- Blank M (1989) Small perturbations of chaotic dynamical systems. (Russian). *Uspekhi Mat Nauk* 44:3–28, 203. Translation in: *Russ Math Surv* 44:1–33
- Blank M (1997) Discreteness and continuity in problems of chaotic dynamics. In: *Translations of mathematical monographs*, vol 161. American Mathematical Society, Providence
- Blank M, Keller G (1997) Stochastic stability versus localization in one-dimensional chaotic dynamical systems. *Nonlinearity* 10:81–107
- Blank M, Keller G, Liverani C (2002) Ruelle–Perron–Frobenius spectrum for Anosov maps. *Nonlinearity* 15:1905–1973
- Bochi J (2002) Genericity of zero Lyapunov exponents. *Ergodic Theory Dynam Syst* 22:1667–1696
- Bochi J, Viana M (2005) The Lyapunov exponents of generic volume-preserving and symplectic maps. *Ann Math* 161(2):1423–1485
- Bolsinov AV, Taimanov IA (2000) Integrable geodesic flows with positive topological entropy. *Invent Math* 140:639–650
- Bonano C, Collet P (2006) Complexity for extended dynamical systems. arXiv:math/0609681
- Bonatti C, Crovisier S (2004) Recurrence et genericité. *Invent Math* 158(1):33–104
- Bonatti C, Viana M (2000) SRB measures for partially hyperbolic systems whose central direction is mostly contracting. *Israel J Math* 115:157–193

- Bonatti C, Viana M (2004) Lyapunov exponents with multiplicity 1 for deterministic products of matrices. *Ergodic Theory Dynam Syst* 24(5):1295–1330
- Bonatti C, Diaz L, Viana M (2005) Dynamics beyond uniform hyperbolicity. A global geometric and probabilistic perspective. In: *Mathematical physics, III. Encyclopedia of mathematical sciences*, vol 102. Springer, Berlin
- Bowen R (1975) Equilibrium states and the ergodic theory of Anosov diffeomorphisms. In: *Lecture notes in mathematics*, vol 470. Springer, Berlin/New York
- Bowen R (1979) Hausdorff dimension of quasi-circles. *Inst Hautes Études Sci Publ Math* 50:11–25
- Bowen R, Ruelle D (1975) Ergodic theory of Axiom A flows. *Invent Math* 29:181–202
- Boyle M, Downarowicz T (2004) The entropy theory of symbolic extensions. *Invent Math* 156:119–161
- Boyle M, Fiebig D, Fiebig U (2002) Residual entropy, conditional entropy and subshift covers. *Forum Math* 14:713–757
- Boyle M, Buzzi J, Gomez R (2006) Ricardo almost isomorphism for countable state Markov shifts. *J Reine Angew Math* 592:23–47
- Brain M, Berger A (2001) Chaos and chance. In: *An introduction to stochastic aspects of dynamics*. Walter de Gruyter, Berlin
- Brin M (2001) Appendix A. In: Barreira L, Pesin Y (eds) *Lectures on Lyapunov exponents and smooth ergodic theory*. *Proc Sympos Pure Math*, 69, Smooth ergodic theory and its applications, Seattle, WA, 1999. American Mathematical Society, Providence, pp 3–106
- Brin M, Stuck G (2002) *Introduction to dynamical systems*. Cambridge University Press, Cambridge
- Broise A (1996a) Transformations dilatantes de l'intervalle et theoremes limites. *Etudes spectrales d'operateurs de transfert et applications*. *Asterisque* 238:1–109
- Broise A (1996b) Transformations dilatantes de l'intervalle et theoremes limites. *Astérisque* 238:1–109
- Bruin H, Keller G, Nowicki T, van Strien S (1996) Wild Cantor attractors exist. *Ann Math* 143(2):97–130
- Buzzi J (1997) Intrinsic ergodicity of smooth interval maps. *Israel J Math* 100:125–161
- Buzzi J (2000a) Absolutely continuous invariant probability measures for arbitrary expanding piecewise $\mathbf{R}$ -analytic mappings of the plane. *Ergodic Theory Dynam Syst* 20:697–708
- Buzzi J (2000b) On entropy-expanding maps. Unpublished
- Buzzi J (2001) No or infinitely many a.c.i.p. for piecewise expanding C^r maps in higher dimensions. *Commun Math Phys* 222:495–501
- Buzzi J (2005) Subshifts of quasi-finite type. *Invent Math* 159:369–406
- Buzzi J (2006) Puzzles of quasi-finite type, zeta functions and symbolic dynamics for multi-dimensional maps. [arXiv:math/0610911](https://arxiv.org/abs/math/0610911)
- Buzzi J. Hyperbolicity through Entropies. Saint-Flour Summer probability school lecture notes (to appear)
- Carleson L, Gamelin TW (1993) *Complex dynamics*. In: *Universitext: tracts in mathematics*. Springer, New York
- Chazottes JR, Gouëzel S (2007) On almost-sure versions of classical limit theorems for dynamical systems. *Probab Theory Relat Fields* 138:195–234
- Chernov N (1999) Statistical properties of piecewise smooth hyperbolic systems in high dimensions. *Discrete Contin Dyn Syst* 5(2):425–448
- Chernov N, Markarian R, Troubetzkoy S (2000) Invariant measures for Anosov maps with small holes. *Ergodic Theory Dynam Syst* 20:1007–1044
- Christensen JRR (1972) On sets of Haar measure zero in abelian Polish groups. *Israel J Math* 13:255–260
- Collet P (1996) Some ergodic properties of maps of the interval. In: *Dynamical systems, Temuco, 1991/1992*. *Travaux en Cours*, 52. Hermann, Paris, pp 55–91
- Collet P, Eckmann JP (2006) *Concepts and results in chaotic dynamics: a short course*. In: *Theoretical and mathematical physics*. Springer, Berlin
- Collet P, Galves A (1995) Asymptotic distribution of entrance times for expanding maps of the interval. In: *Dynamical systems and applications*. *World Sci Ser Appl Anal*, vol 4. World Scientific Publishing, River Edge, pp 139–152
- Collet P, Courbage M, Mertens S, Neishtadt A, Zaslavsky G (eds) (2005) *Chaotic dynamics and transport in classical and quantum systems*. In: *Proceedings of the International Summer School of the NATO Advanced Study Institute held in Cargese, August 18–30, 2003*. Edited by NATO science series II: mathematics, physics and chemistry, 182. Kluwer, Dordrecht
- Contreras G, Lopes AO, Thieullen P (2001) Lyapunov minimizing measures for expanding maps of the circle, English summary. *Ergodic Theory Dynam Syst* 21:1379–1409
- Cowieson WJ (2002) Absolutely continuous invariant measures for most piecewise smooth expanding maps. *Ergodic Theory Dynam Syst* 22:1061–1078
- Cowieson WJ, Young LS (2005) SRB measures as zero-noise limits. *Ergodic Theory Dynam Syst* 25:1115–1138
- Crovisier S (2006a) Birth of homoclinic intersections: a model for the central dynamics of partially hyperbolic systems. <http://www.arxiv.org/abs/math/0605387>
- Crovisier S (2006b) Periodic orbits and chain-transitive sets of C^1 -diffeomorphisms. *Publ Math Inst Hautes Etudes Sci* 104:87–141
- Crovisier S (2006c) Perturbation of C^1 -diffeomorphisms and generic conservative dynamics on surfaces. *Dynamique des diffeomorphismes conservatifs des surfaces: un point de vue topologique*, 21, *Panor syntheses*. Soc Math France, Paris, pp 1–33
- Cvitanović P, Gunaratne GH, Procaccia I (1988) Topological and metric properties of Henon-type strange attractors. *Phys Rev A* 38(3):1503–1520
- de Carvalho A, Hall T (2002) How to prune a horseshoe. *Nonlinearity* 15:R19–R68

- de Castro A (2002) Backward inducing and exponential decay of correlations for partially hyperbolic attractors. *Israel J Math* 130:29–75
- de la Llave R (2001) A tutorial on KAM theory. In: *Smooth ergodic theory and its applications*, Seattle, WA, 1999. *Proc Sympos Pure Math*, vol 69. American Mathematical Society, Providence, pp 175–292
- de Melo W, van Strien S (1993) One-dimensional dynamics. In: *Ergebnisse der Mathematik und ihrer Grenzgebiete* (3), 25. Springer, Berlin
- Denker M, Philipp W (1984) Approximation by Brownian motion for Gibbs measures and flows under a function. *Ergodic Theory Dynam Syst* 4:541–552
- Detmers MF, Liverani C (2007) Stability of statistical properties in two-dimensional piecewise hyperbolic maps. *Trans Am Math Soc* 360:4777–4814
- Devaney RL (1989) An introduction to chaotic dynamical systems, 2nd edn. Addison–Wesley, Redwood
- Diaz L, Rocha J, Viana M (1996) Strange attractors in saddle-node cycles: prevalence and globality. *Invent Math* 125:37–74
- Dolgopyat D (2000) On dynamics of mostly contracting diffeomorphisms. *Commun Math Phys* 213:181–201
- Dolgopyat D (2002) On mixing properties of compact group extensions of hyperbolic systems. *Israel J Math* 130:157–205
- Dolgopyat D (2004a) Limit theorems for partially hyperbolic systems. *Trans Am Math Soc* 356:1637–1689
- Dolgopyat D (2004b) On differentiability of SRB states for partially hyperbolic systems. *Invent Math* 155:389–449
- Downarowicz T (2005) Entropy structure. *J Anal Math* 96:57–116
- Downarowicz T, Newhouse S (2005) Symbolic extensions and smooth dynamical systems. *Invent Math* 160:453–499
- Downarowicz T, Serafin J (2003) Possible entropy functions. *Israel J Math* 135:221–250
- Duhem P (1906) *La théorie physique, son objet et sa structure*. Vrin, Paris, 1981
- Eckmann JP, Ruelle D (1985) Ergodic theory of chaos and strange attractors. *Rev Mod Phys* 57:617–656
- Eskin A, McMullen C (1993) Mixing, counting, and equidistribution in Lie groups. *Duke Math J* 71:181–209
- Fiedler B (ed) (2001) *Ergodic theory, analysis, and efficient simulation of dynamical systems*. Springer, Berlin
- Fiedler B (ed) (2002) *Handbook of dynamical systems*, vol 2. North-Holland, Amsterdam
- Fisher A, Lopes AO (2001) Exact bounds for the polynomial decay of correlation, $1/f$ noise and the CLT for the equilibrium state of a non-Hölder potential. *Nonlinearity* 14:1071–1104
- Furstenberg H (1981) Recurrence in ergodic theory and combinatorial number theory. In: *MB Porter lectures*. Princeton University Press, Princeton
- Gallavotti G (1999) Statistical mechanics. In: *A short treatise. Texts and monographs in physics*. Springer, Berlin
- Gatzouras D, Peres Y (1997) Invariant measures of full dimension for some expanding maps. *Ergodic Theory Dynam Syst* 17:147–167
- Glasner E, Weiss B (1993) Sensitive dependence on initial conditions. *Nonlinearity* 6:1067–1075
- Gordin MI (1969) The central limit theorem for stationary processes. (Russian). *Dokl Akad Nauk SSSR* 188:739–741
- Gouëzel S (2004) Central limit theorem and stable laws for intermittent maps. *Probab Theory Relat Fields* 128:82–122
- Gouëzel S (2005) Berry–Esseen theorem and local limit theorem for non uniformly expanding maps. *Ann Inst H Poincaré* 41:997–1024
- Gouëzel S, Liverani C (2006) Banach spaces adapted to Anosov systems. *Ergodic Theory Dynam Syst* 26:189–217
- Graczyk J, Świątek G (1997) Generic hyperbolicity in the logistic family. *Ann Math* 146(2):1–52
- Gromov M (1987) Entropy, homology and semialgebraic geometry. *Séminaire Bourbaki*, vol 1985/86. *Astérisque* No 145–146:225–240
- Guckenheimer J (1979) Sensitive dependence on initial conditions for one-dimensional maps. *Commun Math Phys* 70:133–160
- Guckenheimer J, Holmes P (1990) *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Revised and corrected reprint of the 1983 original. In: *Applied mathematical sciences*, vol 42. Springer, New York
- Guckenheimer J, Wechselberger M, Young LS (2006) Chaotic attractors of relaxation oscillators. *Nonlinearity* 19:701–720
- Guivarc’h Y, Hardy J (1998) Théorèmes limites pour une classe de chaînes de Markov et applications aux difféomorphismes d’Anosov. *Ann Inst H Poincaré* 24:73–98
- Gurevic BM (1996) Stably recurrent nonnegative matrices. (Russian) *Uspekhi Mat Nauk* 51(3)(309):195–196; translation in: *Russ Math Surv* 51(3):551–552
- Gurevic BM, Savchenko SV (1998) Thermodynamic formalism for symbolic Markov chains with a countable number of states. (Russian). *Uspekhi Mat Nauk* 53(2)(320):3–106; translation in: *Russ Math Surv* 53(2):245–344
- Gutzwiller M (1990) *Chaos in classical and quantum mechanics*. In: *Interdisciplinary applied mathematics*, vol 1. Springer, New York
- Hadamard J (1898) Les surfaces à courbures opposées et leurs lignes géodésiques. *J Math Pures Appl* 4:27–73
- Hasselblatt B, Katok A (eds) (2002, 2006) *Handbook of dynamical systems*, vol 1A and 1B. Elsevier, Amsterdam
- Hasselblatt B, Katok A (2003) *A first course in dynamics*. In: *With a panorama of recent developments*. Cambridge University Press, New York
- Hennion H (1993) Sur un théorème spectral et son application aux noyaux lipschitziens. *Proc Am Math Soc* 118:627–634

- Hénon M (1976) A two-dimensional mapping with a strange attractor. *Commun Math Phys* 50:69–77
- Hesiod (1987) *Theogony*. Focus/R. Pullins, Newburyport
- Hochman M (2006) Upcrossing inequalities for stationary sequences and applications. [arXiv:math/0608311](https://arxiv.org/abs/math/0608311)
- Hofbauer F (1979) On intrinsic ergodicity of piecewise monotonic transformations with positive entropy. *Israel J Math* 34(3):213–237
- Hofbauer F (1985) Periodic points for piecewise monotonic transformations. *Ergodic Theory Dynam Syst* 5: 237–256
- Hofbauer F, Keller G (1982) Ergodic properties of invariant measures for piecewise monotonic transformations. *Math Z* 180:119–140
- Hofbauer F, Keller G (1990) Quadratic maps without asymptotic measure. *Commun Math Phys* 127: 319–337
- Hofer H, Zehnder E (1994) *Symplectic invariants and Hamiltonian dynamics*. Birkhäuser, Basel
- Host B, Kra B (2005) Nonconventional ergodic averages and nilmanifolds. *Ann Math* 161(2):397–488
- Hua Y, Saghin R, Xia Z (2006) Topological entropy and partially hyperbolic diffeomorphisms. [arXiv:math/0608720](https://arxiv.org/abs/math/0608720)
- Hunt FY (1998) Unique ergodicity and the approximation of attractors and their invariant measures using Ulam’s method. (English summary). *Nonlinearity* 11:307–317
- Hunt BR, Sauer T, Yorke JA (1992) Prevalence: a translation-invariant “almost every” on infinite-dimensional spaces. *Bull Am Math Soc (NS)* 27: 217–238
- Ibragimov I, Linnik Y, Kingman JFC (ed) (trans) (1971) Independent and stationary sequences of random variables. Wolters-Noordhoff, Groningen
- Ionescu Tulcea CT, Marinescu G (1950) *Théorie ergodique pour des classes d’opérations non complètement continues*. (French). *Ann Math* 52(2):140–147
- Ishii Y (1997) Towards a kneading theory for Lozi mappings. I. A solution of the pruning front conjecture and the first tangency problem. *Nonlinearity* 10:731–747
- Jakobson MV (1981) Absolutely continuous invariant measures for one-parameter families of one-dimensional maps. *Commun Math Phys* 81:39–88
- Jenkinson O (2007) Optimization and majorization of invariant measures. *Electron Res Announc Am Math Soc* 13:1–12
- Katok A (1980) Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst Hautes Etudes Sci Publ Math* 51:137–173
- Katok A, Hasselblatt B (1995) Introduction to the modern theory of dynamical systems. In: *Encyclopedia of mathematics and its applications*, 54. With a supplementary chapter by: Katok A, Mendoza L. Cambridge University Press, Cambridge
- Keller G (1984) On the rate of convergence to equilibrium in one-dimensional systems. *Commun Math Phys* 96: 181–193
- Keller G (1990) Exponents, attractors and Hopf decompositions for interval maps. *Ergodic Theory Dynam Syst* 10:717–744
- Keller G, Liverani C (1999) Stability of the spectrum for transfer operators. *Ann Scuola Norm Sup Pisa Cl Sci* 28(4):141–152
- Keller G, Nowicki T (1992) Spectral theory, zeta functions and the distribution of periodic points for Collet-Eckmann maps. *Commun Math Phys* 149:31–69
- Kifer Y (1977) Small random perturbations of hyperbolic limit sets. (Russian). *Uspekhi Mat Nauk* 32(1):193–194
- Kifer Y (1986) Ergodic theory of random transformations. In: *Progress in probability and statistics*, vol 10. Birkhäuser Boston, Inc, Boston
- Kifer Y (1988) Random perturbations of dynamical systems. In: *Progress in probability and statistics*, vol 16. Birkhäuser Boston, Inc, Boston
- Kifer Y (1990) Large deviations in dynamical systems and stochastic processes. *Trans Am Math Soc* 321:505–524
- Kifer Y (1997) Computations in dynamical systems via random perturbations. (English summary). *Discrete Contin Dyn Syst* 3:457–476
- Kolyada SF (2004) Li-Yorke sensitivity and other concepts of chaos. *Ukr Math J* 56:1242–1257
- Kozlovski OS (2003) Axiom A maps are dense in the space of unimodal maps in the C^k topology. *Ann Math* 157(2):1–43
- Kozlovski O, Shen W, van Strien S (2007) Density of Axiom A in dimension one. *Ann Math* 166:145–182
- Krengel U (1983) *Ergodic theorems*. De Gruyter, Berlin
- Krzyżewski K, Szlenk W (1969) On invariant measures for expanding differentiable mappings. *Stud Math* 33: 83–92
- Lacroix Y (2002) Possible limit laws for entrance times of an ergodic aperiodic dynamical system. *Israel J Math* 132:253–263
- Ladler RL, Marcus B (1979) Topological entropy and equivalence of dynamical systems. *Memoirs Am Math Soc* 20(219)
- Lastoa A, Yorke J (1973) On the existence of invariant measures for piecewise monotonic transformations. *Trans Am Math Soc* 186:481–488
- Ledrappier F (1984) Propriétés ergodiques des mesures de Sinai. *Inst Hautes Etudes Sci Publ Math* 59:163–188
- Ledrappier F, Young LS (1985) The metric entropy of diffeomorphisms. I Characterization of measures satisfying Pesin’s entropy formula. II Relations between entropy, exponents and dimension. *Ann Math* 122(2): 509–539, 540–574
- Leplaideur R (2004) Existence of SRB-measures for some topologically hyperbolic diffeomorphisms. *Ergodic Theory Dynam Syst* 24:1199–1225
- Li M, Vitanyi P (1997) An introduction to Kolmogorov complexity and its applications. In: *Graduate texts in computer science*, 2nd edn. Springer, New York
- Li TY, Yorke JA (1975) Period three implies chaos. *Am Math Monthly* 82:985–992
- Lind D, Marcus B (1995) *An introduction to symbolic dynamics and coding*. Cambridge University Press, Cambridge
- Liu PD, Qian M, Zhao Y (2003) Large deviations in Axiom A endomorphisms. *Proc R Soc Edinb Sect A* 133: 1379–1388

- Liverani C (1995a) Decay of correlations. *Ann Math* 142: 239–301
- Liverani C (1995b) Decay of correlations in Piecewise expanding maps. *J Stat Phys* 78:1111–1129
- Liverani C (1996) Central limit theorem for deterministic systems. In: Ledrappier F, Lewowicz J, Newhouse S (eds) *International conference on dynamical systems*, Montevideo, 1995. Pitman research notes. Math 362: 56–75
- Liverani C (2001) Rigorous numerical investigation of the statistical properties of piecewise expanding maps – a feasibility study. *Nonlinearity* 14:463–490
- Liverani C, Tsujii M (2006) Zeta functions and dynamical systems. (English summary). *Nonlinearity* 19: 2467–2473
- Lyubich M (1997) Dynamics of quadratic polynomials. I, II. *Acta Math* 178:185–247, 247–297
- Mañé R (1982) An ergodic closing lemma. *Ann Math* 116(2):503–540
- Mañé R (1985) Hyperbolicity, sinks and measures in one-dimensional dynamics. *Commun Math Phys* 100: 495–524
- Mañé R (1988) A proof of the C^1 stability conjecture. *Publ Math IHES* 66:161–210
- Margulis G (2004) On some aspects of the theory of Anosov systems. In: *With a survey by Richard Sharp: periodic orbits of hyperbolic flows*. Springer monographs in mathematics. Springer, Berlin
- Melbourne I, Nicol M (2005) Almost sure invariance principle for nonuniformly hyperbolic systems. *Commun Math Phys* 260(1):131–146
- Milnor J, Thurston W (1988) On iterated maps of the interval. In: *Dynamical systems*, College Park, MD, 1986–87. *Lecture Notes in Math*, vol 1342. Springer, Berlin, pp 465–563
- Misiurewicz M (1976a) A short proof of the variational principle for a Z^n_+ action on a compact space. *Bull Acad Polon Sci Sér Sci Math Astronom Phys* 24(12): 1069–1075
- Misiurewicz M (1976b) Topological conditional entropy. *Stud Math* 55:175–200
- Misiurewicz M (1979) Horseshoes for mappings of the interval. *Bull Acad Polon Sci Sér Sci Math* 27:167–169
- Misiurewicz M (1995) Continuity of entropy revisited. In: *Dynamical systems and applications*. World Sci Ser Appl Anal, vol 4. World Scientific Publishing, River Edge, pp 495–503
- Misiurewicz M, Smítal J (1988) Smooth chaotic maps with zero topological entropy. *Ergodic Theory Dynam Syst* 8:421–424
- Mora L, Viana M (1993) Abundance of strange attractors. *Acta Math* 171:1–71
- Moreira CG, Palis J, Viana M (2001) Homoclinic tangencies and fractal invariants in arbitrary dimension. *CRAS* 333:475–480
- Murray JD (2002, 2003) *Mathematical biology*. I and II An introduction, 3rd edn. In: *Interdisciplinary applied mathematics*, 17 and 18. Springer, New York
- Nagaev SV (1957) Some limit theorems for stationary Markov chains. *Theor Probab Appl* 2:378–406
- Newhouse SE (1974) Diffeomorphisms with infinitely many sinks. *Topology* 13:9–18
- Newhouse SE (1989) Continuity properties of entropy. *Ann Math* 129(2):215–235; Erratum: *Ann of Math* 131(2):409–410
- Nussbaum RD (1970) The radius of the essential spectrum. *Duke Math J* 37:473–478
- Ornstein D (1970) Bernoulli shifts with the same entropy are isomorphic. *Adv Math* 4:337–352
- Ornstein D, Weiss B (1988) On the Bernoulli nature of systems with some hyperbolic structure. *Ergodic Theory Dynam Syst* 18:441–456
- Ovid (2005) *Metamorphosis*. W.W. Norton, New York
- Palis J (2000) A global view of dynamics and a conjecture on the senseness of tinitude of attractors. *Asterisque* 261:335–347
- Palis J, Yoccoz JC (2001) Fers a cheval non-uniformement hyperboliques engendres par une bifurcation homocline et densite nulle des attracteurs. *CRAS* 333:867–871
- Pesin YB (1976) Families of invariant manifolds corresponding to non-zero characteristic exponents. *Math USSR Izv* 10:1261–1302
- Pesin YB (1977) Characteristic exponents and smooth ergodic theory. *Russ Math Surv* 324:55–114
- Pesin Ya (1997) Dimension theory in dynamical systems. In: *Contemporary views and applications*. Chicago lectures in mathematics. University of Chicago Press, Chicago
- Pesin Ya, Sinai Ya (1982) Gibbs measures for partially hyperbolic attractors. *Ergodic Theory Dynam Syst* 2: 417–438
- Piorek J (1985) On the generic chaos in dynamical systems. *Univ Iagel Acta Math* 25:293–298
- Plykin RV (2002) On the problem of the topological classification of strange attractors of dynamical systems. *Uspekhi Mat Nauk* 57:123–166. Translation in: *Russ Math Surv* 57:1163–1205
- Poincare H (1892) *Les methodes nouvelles de la mecanique céleste*. Gauthier–Villars, Paris
- Pollicott M, Sharp R (2002) Invariance principles for interval maps with an indifferent fixed point. *Commun Math Phys* 229:337–346
- Przytycki F (1977) On U -stability and structural stability of endomorphisms satisfying. *Axiom A Studia Math* 60: 61–77
- Pugh C, Shub M (1999) Ergodic attractors. *Trans Am Math Soc* 312:1–54
- Pujals E, Rodriguez-Hertz F (2007) Critical points for surface diffeomorphisms. *J Mod Dyn* 1:615–648
- Puu T (2000) Attractors, bifurcations, and chaos. In: *Nonlinear phenomena in economics*. Springer, Berlin
- Rees M (1981) A minimal positive entropy homeomorphism of the 2-torus. *J Lond Math Soc* 23(2):537–550
- Robinson RC (2004) *An introduction to dynamical systems: continuous and discrete*. Pearson Prentice Hall, Upper Saddle River

- Rousseau-Egele J (1983) Un théorème de la limite locale pour une classe de transformations dilatantes et monotones par morceaux. *Ann Probab* 11:772–788
- Ruelle D (1968) Statistical mechanics of a one-dimensional lattice gas. *Commun Math Phys* 9:267–278
- Ruelle D (1976) A measure associated with axiom-A attractors. *Am J Math* 98:619–654
- Ruelle D (1978) An inequality for the entropy of differentiable maps. *Bol Soc Brasil Mat* 9:83–87
- Ruelle D (1982) Repellers for real analytic maps. *Ergodic Theory Dynam Syst* 2:99–107
- Ruelle D (1989) The thermodynamic formalism for expanding maps. *Commun Math Phys* 125:239–262
- Ruelle D (2004) Thermodynamic formalism. In: *The mathematical structures of equilibrium statistical mechanics*, 2nd edn. Cambridge Mathematical Library, Cambridge University Press, Cambridge
- Ruelle D (2005) Differentiating the absolutely continuous invariant measure of an interval map f with respect to f . *Commun Math Phys* 258:445–453
- Ruette S (2003a) On the Vere-Jones classification and existence of maximal measures for countable topological Markov chains. *Pacific J Math* 209:366–380
- Ruette S (2003b) Chaos on the interval. <http://www.math.u-psud.fr/%7Eruette>
- Saari DG (2005) Collisions, rings, and other Newtonian N -body problems. In: *CBMS regional conference series in mathematics*, 104. Published for the Conference Board of the Mathematical Sciences, Washington, DC. American Mathematical Society, Providence
- Saperstone SH, Yorke JA (1971) Controllability of linear oscillatory systems using positive controls. *SIAM J Control* 9:253–262
- Sarig O (1999) Thermodynamic formalism for countable Markov shifts. *Ergodic Theory Dynam Syst* 19:1565–1593
- Sarig O (2001) Phase transitions for countable topological Markov shifts. *Commun Math Phys* 217:555–577
- Sataev EA (1992) Invariant measures for hyperbolic mappings with singularities. *Uspekhi Mat Nauk* 47:147–202. Translation in: *Russ Math Surv* 47:191–251
- Saussol B (2000) Absolutely continuous invariant measures for multidimensional expanding maps. *Israel J Math* 116:223–248
- Saussol B (2006) Recurrence rate in rapidly mixing dynamical systems. *Discrete Contin Dyn Syst* 15(1):259–267
- Shub M (1987) *Global stability of dynamical systems*. Springer, New York
- Simon R (1972) A 3-dimensional Abraham-Smale example. *Proc Am Math Soc* 34:629–630
- Starkov AN (2000) Dynamical systems on homogeneous spaces. In: *Translations of mathematical monographs*, vol 190. American Mathematical Society, Providence
- Stewart P (1964) *Jacobellis v Ohio*. US Rep 378:184
- Szász D (ed) (2000) *Hard ball systems and the Lorentz gas*. In: *Encyclopedia of mathematical sciences*, 101. Mathematical physics, II. Springer, Berlin
- Tsujii M (1992) A measure on the space of smooth mappings and dynamical system theory. *J Math Soc Jpn* 44:415–425
- Tsujii M (2000a) Absolutely continuous invariant measures for piecewise real-analytic expanding maps on the plane. *Commun Math Phys* 208:605–622
- Tsujii M (2000b) Piecewise expanding maps on the plane with singular ergodic properties. *Ergodic Theory Dynam Syst* 20:1851–1857
- Tsujii M (2001) Absolutely continuous invariant measures for expanding piecewise linear maps. *Invent Math* 143:349–373
- Tsujii M (2005) Physical measures for partially hyperbolic surface endomorphisms. *Acta Math* 194:37–132
- Tucker W (1999) The Lorenz attractor exists. *C R Acad Sci Paris Ser I Math* 328:1197–1202
- Vallée B (2006) Euclidean dynamics. *Discrete Contin Dyn Syst* 15:281–352
- van Strien S, Vargas E (2004) Real bounds, ergodicity and negative Schwarzian for multimodal maps. *J Am Math Soc* 17:749–782
- Vasquez CH (2007) Statistical stability for diffeomorphisms with dominated splitting. *Ergodic Theory Dynam Syst* 27:253–283
- Viana M (1993) Strange attractors in higher dimensions. *Bol Soc Bras Mat (NS)* 24:13–62
- Viana M (1997a) Multidimensional nonhyperbolic attractors. *Inst Hautes Etudes Sci Publ Math No* 85:63–96
- Viana M (1997b) Stochastic dynamics of deterministic systems. In: *Lecture notes 21st Braz Math Colloq IMPA*. Rio de Janeiro
- Viana M (1998) Dynamics: a probabilistic and geometric perspective. In: *Proceedings of the International Congress of Mathematicians*, vol I, Berlin, 1998. *Doc Math Extra I*:557–578
- Wang Q, Young LS (2003) Strange attractors in periodically-kicked limit cycles and Hopf bifurcations. *Commun Math Phys* 240:509–529
- Weiss B (2002) *Single orbit dynamics*. American Mathematical Society, Providence
- Ya S (1972) Gibbs measures in ergodic theory. *Uspekhi Mat Nauk* 27(166):21–64
- Yomdin Y (1987) Volume growth and entropy. *Israel J Math* 57:285–300
- Yoshihara KI (2004) Weakly dependent stochastic sequences and their applications. In: *Recent topics on weak and strong limit theorems*, vol XIV. Sanseido Co Ltd, Chiyoda
- Young LS (1982) Dimension, entropy and Lyapunov exponents. *Ergodic Theory Dynam Syst* 6:311–319
- Young LS (1985) Bowen–Ruelle measures for certain piecewise hyperbolic maps. *Trans Am Math Soc* 287:41–48

- Young LS (1986) Stochastic stability of hyperbolic attractors. *Ergodic Theory Dynam Syst* 6:311–319
- Young LS (1990) Large deviations in dynamical systems. *Trans Am Math Soc* 318:525–543
- Young LS (1992) Decay of correlations for certain quadratic maps. *Commun Math Phys* 146:123–138
- Young LS (1995) Ergodic theory of differentiable dynamical systems. *Real and complex dynamical systems*. Hillerad, 1993. In: NATO Adv Sci Inst Ser C Math Phys Sci, vol 464. Kluwer, Dordrecht, pp 293–336
- Young LS (1998) Statistical properties of dynamical systems with some hyperbolicity. *Ann Math* 585–650
- Young LS (1999a) Recurrence times and rates of mixing. *Israel J Math* 110:153–188
- Young LS, Wang D (2001) Strange attractors with one direction of instability. *Commun Math Phys* 218:1–97
- Young LS, Wang D (2006) Nonuniformly expanding 1D maps. *Commun Math Phys* 264:255–282
- Young LS, Wang D (2008) Toward a theory of rank one attractors. *Ann Math* 167:349–480
- Zhang Y (1997) Dynamical upper bounds for Hausdorff dimension of invariant sets. *Ergodic Theory Dynam Syst* 17:739–756

Books and Reviews

- Baladi V (2000b) The magnet and the butterfly: thermodynamic formalism and the ergodic theory of chaotic dynamics. (English summary). In: *Development of mathematics 1950–2000*. Birkhäuser, Basel, pp 97–133
- Bonatti C (2003) Dynamiques generiques: hyperbolicite et transitivite. In: *Seminaire Bourbaki vol 2001/2002*. Asterisque No 290. pp 225–242
- Kuksin SB (2006) Randomly forced nonlinear PDEs and statistical hydrodynamics in 2 space dimensions. In: *Zurich lectures in advanced mathematics*. European Mathematical Society (EMS), Zürich
- Liu PD, Qian M (1995) Smooth ergodic theory of random dynamical systems. In: *Lecture notes in mathematics*, vol 1606. Springer, Berlin
- Ornstein D, Weiss B (1991) Statistical properties of chaotic systems. With an appendix by David Fried. *Bull Am Math Soc (NS)* 24:11–116
- Young LS (1999b) Ergodic theory of chaotic dynamical systems. In: *XIIth International Congress of Mathematical Physics, ICMP '97, Brisbane*. Int Press, Cambridge, MA, pp 131–143



Ergodic Theory: Fractal Geometry

Jörg Schmeling
Center for Mathematical Sciences, Lund
University, Lund, Sweden

is to study those objects. One of the main tools is the fractal dimension theory that helps to extract important properties of geometrically “irregular” sets.

Article Outline

Glossary
Definition of the Subject
Introduction
Preliminaries
Brief Tour Through Some Examples
Dimension Theory of Low-Dimensional
Dynamical Systems – Young’s Dimension
Formula
Dimension Theory of Higher-Dimensional
Dynamical Systems
Hyperbolic Measures
General Theory
Endomorphisms
Multifractal Analysis
Future Directions
Bibliography

Glossary

Dynamical system A (discrete time) dynamical system describes the time evolution of a point in phase space. More precisely a space X is given and the time evolution is given by a map $T: X \rightarrow X$. The main interest is to describe the asymptotic behavior of the trajectories (orbits) $T^n(x)$, i.e. the evolution of an initial point $x \in X$ under the iterates of the map T . More generally one is interested in obtaining information on the geometrically complicated invariant sets or measures which describe the asymptotic behavior.

Fractal geometry Many objects of interest (invariant sets, invariant measures etc.) exhibit a complicated structure that is far from being smooth or regular. The aim of fractal geometry

Definition of the Subject

The connection between fractal geometry and dynamical system theory is very diverse. There is no unified approach and many of the ideas arose from significant examples. Also the dynamical system theory has been shown to have a strong impact on classical fractal geometry. In this article there are first presented some examples showing nontrivial results coming from the application of dimension theory. Some of these examples require a deeper knowledge of the theory of smooth dynamical systems then can be provided here. Nevertheless, the flavor of these examples can be understood. Then there is a brief overview of some of the most developed parts of the application of fractal geometry to dynamical system theory. Of course a rigorous and complete treatment of the theory cannot be given. The cautious reader may wish to check the original papers. Finally, there is an outlook over the most recent developments. This article is by no means meant to be complete. It is intended to give some of the ideas and results from this field.

Introduction

In this section some of the aspects of fractal geometry in dynamical systems are pointed out. Some notions that are used will be defined later on and I intend only to give a flavor of the applications. The nonfamiliar reader will find the definitions in the corresponding sections and can return to this section later. The *geometry* of many invariant sets or invariant measures of dynamical systems (including attractors, measures defining the statistical properties) look very complicated at *all scales*, and their geometry is impossible to

describe using standard geometric tools. For some important classes of dynamical systems, these complicated structures are intensively studied using notions of dimension. In many cases it becomes possible to relate these notions of dimension to other fundamental dynamical characteristics, such as Lyapunov exponents, entropies, pressure, etc.

On the other hand tools from dynamical systems, especially from ergodic theory and thermodynamic formalism, are extremely useful to explore the fractal properties of the objects in question. This includes dimensions of limit sets of geometric constructions (the standard Cantor set being the most famous example), which a priori, are not related to dynamical systems (Furstenberg 1967; Pesin and Weiss 1996). Many dimension formulas for asymptotic sets of dynamical systems are obtained by means of Bowen-type formulas, i.e. as roots of some functionals arising from thermodynamic formalism.

The dimension of a set is a subtle characteristic which measures the *geometric complexity* of the set at arbitrarily fine scales. There are many notions of dimension, and most definitions involve a measurement of geometric complexity at scale ε (which ignores the irregularities of the set at size less than ε) and then considers the limiting measurement as $\varepsilon \rightarrow 0$. A priori (and in general) these different notions can be different. An important result is the affirmative solution of the Eckmann–Ruelle conjecture by Barreira, Pesin and Schmeling (1999), which says that for smooth nonuniformly hyperbolic systems, the pointwise dimension is almost everywhere constant with respect to a hyperbolic measure. This result implies that many dimension characteristics for the measure coincide.

The deep connection between dynamical systems and dimension theory seems to have been first discovered by Billingsley (1978) through several problems in number theory.

Another link between dynamical systems and dimension theory is through Pesin's theory of dimension-like characteristics. This general theory is a unification of many notions of dimension along with many fundamental quantities in dynamical system such as entropies, pressure, etc.

However, there are numerous examples of dynamical systems exhibiting pathological behavior with respect to fractal geometrical characteristics. In particular higher-dimensional systems seem to be as complicated as general objects considered in geometric measure theory. Therefore, a clean and unified theory is still not available.

The study of characteristic notions like entropy, exponents or dimensions is an essential issue in the theory of dynamical systems. In many cases it helps to classify or to understand the dynamics. Most of these characteristics were introduced for different questions and concepts. For example, entropy was introduced to distinguish non-isomorphic systems and appeared to be a complete invariant for Bernoulli systems (Ornstein). Later, the thermodynamic formalism (see (Ruelle 1978)) introduced new quantities like the pressure. Bowen (1979) and also Ruelle discovered a remarkable connection between the thermodynamic formalism and the dimension theory for invariant sets. Since then many efforts were taken to find the relations between all these different quantities. It occurred that the dimension of invariant sets or measures carries lots of information about the system, combining its combinatorial complexity with its geometric complexity. Unfortunately it is extremely difficult to compute the dimension in general. The general flavor is that local divergence of orbits and global recurrence cause complicated global behavior (chaos). It is impossible to study the exact (infinite) trajectory of all orbits. One way out is to study the statistical properties of "typical" orbits by means of an invariant measure. Although the underlying system might be smooth the invariant measures may often be singular.

Preliminaries

Throughout the article the following situation is considered. Let M be a compact Riemannian manifold without boundary. On M is acting a dynamical system generated by a $C^{1+\alpha}$ diffeomorphism $T: M \rightarrow M$. The presence of a dynamical system provides several important additional tools and methods for the theory of fractal dimensions. Also the theory of fractal dimensions allows one

to draw deep conclusions about the dynamical system. The importance and relevance of the study of fractal dimension will be explained in later sections.

In the next sections some of the most important tools in the fractal theory of dynamical systems are considered. The definitions given here are not necessarily the original definitions but rather the ones which are closer to contemporary use. More details can be found in (Pesin 1997b).

Some Ergodic Theory

Ergodic theory is a powerful method to analyze statistical properties of dynamical systems. All the following facts can be found in standard books on ergodic theory like (Petersen 1983; Walters 1982).

The main idea in ergodic theory is to relate global quantities to observations along single orbits. Let us consider an **invariant measure**: $\mu(f^{-1}A) = \mu(A)$ for all measurable sets A . Such a measure “selects” typical trajectories. It is important to note that the properties vary with the invariant measures. Any such invariant measure can be decomposed into elementary parts (ergodic components).

An invariant measure is called **ergodic** if for any invariant set $A = T^{-1}A$ one has $\mu(A)\mu(M \setminus A) = 0$ (with the agreement $0 \cdot \infty = 0$), i.e. from the measure-theoretic point of view there are no non-trivial invariant subsets.

The importance of ergodic **probability** measures (i.e. $\mu(M) = 1$) lies in the following theorem of Birkhoff.

Theorem 1 (Birkhoff) *Let μ be an ergodic probability measure and $\varphi \in L^1(\mu)$. Then*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(T^k x) = \int_M \varphi d\mu \quad \mu - \text{a.e.}$$

Hausdorff Dimension

With $Z \subset \mathbb{R}^N$ and $s \geq 0$ one defines

$$m_H(s, Z) = \liminf_{\delta \rightarrow 0} \inf_{\{B_i\}} \left\{ \sum_i \text{diam}(B_i)^s : \sup_i \text{diam}(B_i) < \delta \text{ and } \bigcup_i B_i \supset Z \right\}.$$

Note that this limit exists. $m_H(s, Z)$ is called the **s-dimensional outer Hausdorff measure of Z**. It is immediate that there exists a unique value s^* , called the **Hausdorff dimension of Z**, at which $m_H(s, Z)$ jumps from ∞ to 0.

In general it is very hard to find optimal coverings and hence it is often impossible to compute the Hausdorff dimension of a set. Therefore a simpler notion – the lower and upper **Box dimension** – was introduced. The difference to the Hausdorff dimension is that the covering balls are assumed to have the same radius ε . Since then the limit as $\varepsilon \rightarrow 0$ does not have to exist one arrives at the notion of the upper and lower dimension.

Dimension of a Measure

Definition 1 Let $Z \subset \mathbb{R}^N$ and let μ be a probability measure supported on Z . Define the **Hausdorff dimension of the measure μ** by

$$\dim_H(\mu) \equiv \inf_{K \subset Z: \mu(K)=1} \dim_H(K).$$

Pointwise Dimension

Most invariant sets or measures are not strongly self-similar, i.e. the local geometry at arbitrarily fine scales might look different from point to point. Therefore, the notion of pointwise dimension with respect to a Borel probability measure is defined.

Let μ be a Borel probability measure. By $B(x, \varepsilon)$ the ball with center x and radius ε is denoted. The **pointwise dimension** of the measure μ at the point x is defined as

$$d_\mu(x) := \lim_{\varepsilon \rightarrow 0} \frac{\log \mu(B(x, \varepsilon))}{\log \varepsilon}$$

if the above limit exists. If $d_\mu(x) = d$, then for small ε the measure of small balls scales as $\mu(B(x, \varepsilon)) \asymp \varepsilon^d$.

Proposition 1 Suppose μ is a probability measure supported on $Z \subset \mathbb{R}^N$. Then $d_\mu(x) = d$ for μ -almost all $x \in Z$ implies $\dim_H(\mu) = d$.

One should not take the existence of a local dimension (even for *good* measures) for granted. Later on it will be seen that the spectrum of the pointwise dimension (dimension spectrum) is a main object of study in classical multifractal analysis.

The dimension of a measure or a set of measures is its geometric complexity. However, under the presence of a dynamical system one also wants to measure the dynamical (combinatorial) complexity of the system. This leads to the notion of entropy.

Dimension-Like Characteristics and Topological Entropy

Pesin's theory of dimension-like characteristics provides a unified treatment of dimensions and important dynamical quantities like entropies and pressure. The topological entropy of a continuous map f with respect to a subset Z in a metric space (X, ρ) (in particular $X = M$ – a Riemannian manifold) can be defined as a dimension-like characteristic. For each $n \in \mathbb{N}$ and $\varepsilon > 0$, define the **Bowen ball** $B_n(x, \varepsilon) = \{y \in X : \rho(T^i(x), T^i(y)) \leq \varepsilon \text{ for } 0 \leq i \leq n\}$. Then let

$$m_h(Z, \alpha, \varepsilon, n) := \lim_{n \rightarrow \infty} \inf \left\{ \sum_i e^{\alpha n_i} : n_i > n, \bigcup_i B_{n_i}(x, \varepsilon) \supset Z \right\}.$$

This gives rise to an outer measure that jumps from ∞ to 0 at some value α^* . This threshold value α^* is called the **topological entropy of Z** (at scale ε). However, in many situations this value does not depend on ε . The topological entropy is denoted by $h_{\text{top}}(T|Z)$.

If Z is f -invariant and compact, this definition of topological entropy coincides with the usual definition of topological entropy (Walters 1982).

The **entropy of a measure μ** is defined as $h_\mu = \inf \{h_{\text{top}}(T|Z) : \mu(Z) = 1\}$. For ergodic measures this definition coincides with the Kolmogorov–Sinai entropy (see (Pesin 1997b)).

One has to note that in the definition of entropy metric (“round”) balls are substituted by Bowen balls and the metric diameter by the “depth” of the Bowen ball. Therefore, the relation between entropy and dimension is determined by the

relation of “round” balls to “oval” dynamical Bowen balls. If one understands how metric balls can be *efficiently* used to cover dynamical balls one can use the dynamical and relatively easy relation to compute notion of entropy to determine the dimension. However, in higher dimensions this relation is by far nontrivial. A heuristic argument for comparing “round” balls with dynamical balls is given in section “The Kaplan–Yorke Conjecture”.

The Pressure Functional

A useful tool in the dimension analysis of dynamical systems is the pressure functional. It was originally defined by means of statistical physics (thermodynamic formalism) as the free energy (or pressure) of a potential φ (see for example (Ruelle 1978)). However, in this article a dimension-like definition (see (Pesin 1997b)) is more suitable. Again an outer measure using Bowen balls will be used. Let $\varphi : M \rightarrow \mathbb{R}$ be a continuous function and

$$m_P(Z, \alpha, \varepsilon, n, \varphi) = \lim_{n \rightarrow \infty} \inf \left\{ \sum_i \exp(-\alpha n_i) + \sup_{x \in B_{n_i}(x, \varepsilon)} \sum_{k=0}^n \varphi(T^k x) \right\}.$$

This defines an outer measure that jumps from ∞ to 0 as α increases. The threshold value α^* is called the **topological pressure** of the potential φ denoted by $P(\varphi)$. In many situations it does not depend on ε .

There is also a third way of defining the pressure in terms of a variational principle (see (Walters 1982)):

$$P(\varphi) = \sup_{\mu \text{-invariant}} \left(h_\mu + \int_M \varphi \, d\mu \right)$$

Brief Tour Through Some Examples

Before describing the fractal theory of dynamical systems in more detail some ideas about its role are presented. The application of dimension theory has many different aspects. At this point some

(but by far not all) important examples are considered that should give the reader some feeling about the importance and wide use of dimension theory in dynamical system theory.

Dimension of Conformal Repellers: Ruelle's Pressure Formula

Computing or estimating a dimension via a *pressure formula* is a fundamental technique. Explicit properties of pressure help to analyze subtle characteristics. For example, the smooth dependence of the Hausdorff dimension of basic sets for Axiom- A surface diffeomorphisms on the derivative of the map follows from smoothness of pressure.

Ruelle proved the following pressure formula for the Hausdorff dimension of a conformal repeller. A conformal repeller J is an invariant set $T(J) = J = \{x \in M : f^n x \in V \forall n \in \mathbb{N} \text{ and some neighborhood } V \text{ of } J\}$ such that for any $x \in J$ the differential $D_x T = a(x) \text{Iso}_x$ where $a(x)$ is a scalar with $|a(x)| > 1$ and Iso_x an isometry of the tangent space $T_x M$.

Theorem 2 (Ruelle 1982) *Let $T : J \rightarrow J$ be a conformal repeller, and consider the function $t \rightarrow P(-t \log |D_x T|)$, where P denotes pressure. Then*

$$P(-s \log |D_x T|) = 0 \Leftrightarrow s = \dim_H(J)$$

Iterated Function Systems

In fractal geometry one of the most-studied objects are iterated function systems, where there are given n contractions $F_1, \dots, F_n$ of $\mathbb{R}^d$. The unique compact set J fulfilling $J = \cup_i F_i(J)$ is called the attractor of the iterated function system (see (Falconer 1990)). One question is to evaluate its Hausdorff dimension. Often (for example under the open set condition) the attractor J can be represented as the repeller of a piecewise expanding map $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ where the F_i are the inverse branches of the map T . In general it is by far not trivial to determine the dimension of J or even the dimension of a measure sitting on J . The

following example explains some of those difficulties.

Let $1/2 < \lambda < 1$ and consider the maps $F_i : [0, 1] \rightarrow [0, 1]$ given by $F_1(x) = \lambda x$ and $F_2(x) = \lambda x + (1 - \lambda)$. Then the images of F_1, F_2 have an essential overlap and $J = [0, 1]$. If one randomizes this construction in the way that one applies both maps each with probability $1/2$ a probability measure is induced on J . This measure might be absolute continuous with respect to Lebesgue measure or not. Already Erdős realized that for some special values of λ (for example for the inverse of the golden mean) the induced measure is singular. In a breakthrough paper B. Solomyak (1995) proved that for a.e. λ the induced measure is absolutely continuous. A main ingredient in the proof is a transversality condition in the parameter space: the images of arbitrary two random samples of the (infinite) applications of the maps F_i have to cross with nonzero speed when the parameter λ changes. This is a general mechanism which allows one to handle more general situations.

Homoclinic Bifurcations for Dissipative Surface Diffeomorphisms

Homoclinic tangencies and their bifurcations play a fundamental role in the theory of dynamical systems (Palis and Takens 1987, 1993, 1994). Systems with homoclinic tangencies have a complicated and subtle quasi-local behavior. Newhouse showed that homoclinic tangencies can persist under small perturbations, and that horseshoes may co-exist with infinitely many sinks in a neighborhood of the homoclinic orbit and hence the system is not hyperbolic (**Newhouse phenomenon**).

Let $T_\mu : M^2 \rightarrow M^2$ be a smooth parameter family of surface diffeomorphisms that exhibits for $\mu = 0$ an invariant hyperbolic set Λ_0 (horseshoe) and undergoes a homoclinic bifurcation. The Hausdorff dimension of the hyperbolic set Λ_0 for T_0 determines whether hyperbolicity is the *typical* dynamical phenomenon near T_0 or not. If $\dim_H \Lambda_0 < 1$, then hyperbolicity is the prevalent dynamical phenomenon near f_0 . This is not the case if $\dim_H \Lambda_0 > 1$.

More precisely, let $\mathcal{NW}_\mu$ denote the set of nonwandering points of T_μ in an open neighborhood of Λ_0 after the homoclinic bifurcation. Let ℓ denote Lebesgue measure.

Theorem 3 (Palis and Takens 1987, 1993) *If $\dim_H \Lambda_0 < 1$, then*

$$\lim_{\mu_0 \rightarrow 0} \frac{\ell\{\mu \in [0, \mu_0] : \mathcal{NW}_\mu \text{ is hyperbolic}\}}{\mu_0} = 1.$$

The hyperbolicity co-exists with the Newhouse phenomena for a residual set of parameter values.

Theorem 4 (Palis–Yoccoz) *If $\dim_H \Lambda_0 > 1$, then*

$$\lim_{\mu_0 \rightarrow 0} \frac{\ell\{\mu \in [0, \mu_0] : \mathcal{NW}_\mu \text{ is hyperbolic}\}}{\mu_0} < 1.$$

Some Applications to Number Theory

Sometimes dimension problems in number theory can be transferred to dynamical systems and attacked using tools from dynamics.

Example Diadic expansion of numbers: Consider a real number expanded in base 2, i.e. $x = \sum_{n=1}^{\infty} x_n/2^n$. Let

$$X_p = \left\{ x : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} x_k = p \right\}.$$

Borel showed that $\ell(X_{1/2}) = 1$, where ℓ denotes Lebesgue measure. This result is an easy consequence of the Birkhoff ergodic theorem applied to the characteristic function of the digit 1 (which in this simple case is the Strong Law of Large Numbers for i.i.d. processes).

One can ask how large is the set X_p in general. Eggleston (1952) discovered the following wonderful dimension formula, which Billingsley (1978) interpreted in terms of dynamics and reproved using tools from ergodic theory.

Theorem 5 (Eggleston 1952) *The Hausdorff dimension of X_p is given by*

$$\dim_H(X_p) = (1/\log 2)[-p \log p - (1-p) \log(1-p)].$$

The underlying dynamical system is $E_2(x) = 2x \mod 1$ and the dimension is the dimension of the Bernoulli p measure. Other cases included that by Rényi who proposed generalization of the base d expansion from integer base d to noninteger base β . In this case the underlying dynamical system is the (in general non-Markovian) beta shift $\beta(x) = \beta x \mod 1$ studied in (Schmeling 1997).

There are many investigations concerning the approximation of real numbers by rationals by using dynamical methods. The underlying dynamical system in this case is the Gauss map $T(x) = (1/x) \mod 1$. This map is uniformly expanding, but has infinitely many branches.

Example Continued fraction approximation of numbers

Consider the continued fraction expansion of $x \in [0, 1]$, i.e.,

$$x = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \ddots}}}} = [a_1, a_2, a_3, a_4, \dots],$$

and the approximants $p_n(x)/q_n(x) = [a_1, a_2, \dots, a_n]$ (i.e. the approximation given by the finite continued fraction after step n).

The set of numbers which admit a *faster approximation* by rational numbers is defined as follows. For $\tau \geq 2$, let $\mathcal{F}_\tau$

$$\mathcal{F}_\tau = \left\{ x \in [0, 1] : \left| x - \frac{p}{q} \right| \leq \frac{1}{q^\tau} \text{ for infinitely many } p/q \right\}.$$

It is well known that this set has zero measure for each $\tau \geq 2$. Jarník (1931) computed the Hausdorff dimension of $\mathcal{F}_\tau$ and showed that $\dim_H(\mathcal{F}_\tau) = 2/\tau$. However, nowadays there are methods from dynamical systems which not only allow unified proofs but also can handle more subtle subsets of the reals defined by some

properties of their continued fraction expansion (see for example (Aihua et al. 2005; Pollicott and Weiss 1999)).

Infinite Iterated Function Systems and Parabolic Systems

In the previous section a system with infinitely many branches appeared. This can be regarded as an iterated function system with infinitely many maps F_i . This situation is quite general. If one considers a (one-dimensional) system with a parabolic (indifferent) fixed point, i.e. there is a fixed point where the derivative has absolute value equal to 1, one often uses an induced system. For this one chooses nearby the parabolic point a higher iterate of the map in order to achieve uniform expansion away from the parabolic point. This leads to infinitely many branches since the number of iterates has to be increased the closer the parabolic point is. The main difference to the finite iterated function system is that the setting is no longer compact and many properties of the pressure functional are lost.

Mauldin and Urbański and others (see for example (Aihua et al. 2005; Mauldin and Urbański 1996, 2000, 2002)) developed a thermodynamic formalism adapted to the pressure functional for infinite iterated function systems. Besides noncompactness one of the main problems is the loss of analyticity (phase transitions) and convexity of the pressure functional for infinite iterated function systems.

Complex Dynamics

Let $T: \mathbb{C} \rightarrow \mathbb{C}$ be a polynomial and J its Julia set (repeller of this system). If this set is hyperbolic, i.e. the derivative at each point has absolute value larger than 1 the study of the dimension can be related to the study of a finite iterated function system. However, in the presence of a parabolic fixed point this leads to an infinite iterated function system.

If one considers the coefficients of the polynomial as parameters one often sees a qualitative change in the asymptotic behavior. For example, the classical Mandelbrot set for polynomials $z^2 + c$ is the locus of values $c \in \mathbb{C}$ for which the orbit of the origin stays bounded. This set is well

known to be fractal. However, its complete description is still not available.

Embedology and Computational Aspects of Dimension

Tools from dynamical systems are becoming increasingly important to study the time evolution of deterministic systems in engineering and the physical and biological sciences. One of the main ideas is to model a “real world” system by a smooth dynamical system which possesses a strange attractor with a natural ergodic invariant measure. When studying a complicated real world system, one can measure only a very small number of variables. The challenge is to reconstruct the underlying attractor from the time measurement of a scalar quantity. An idealized measurement is considered as a function $h: M^n \rightarrow \mathbb{R}$.

The main tool researchers currently use to *reconstruct* the model system is called attractor reconstruction (see papers and references in (Ott et al. 1994)). This method is based on embedding with time delays (see the influential paper (Crutchfield et al. 1980), where the authors attribute the idea of delay coordinates to Ruelle), where one attempts to reconstruct the attractor for the model using a single long trajectory. Then one considers the points in $\mathbb{R}^{p+1}$ defined by $(x_k, x_{k+\tau}, x_{k+2\tau}, \dots, x_{k+p\tau})$.

Takens (1981) showed that for a smooth $T: M^n \rightarrow M^n$ and for typical smooth h , the mapping $\varphi: M^n \rightarrow \mathbb{R}^{2n+1}$ defined by $x \rightarrow (h(x), h(f(x)), \dots, h(f^{2n\tau}(x)))$ is an embedding. Since the box dimension of the attractor Λ may be much less than the dimension of the ambient manifold n , an interesting mathematical question is whether there exists $p < 2n + 1$ such that the mapping on the attractor $\varphi: \Lambda \rightarrow \mathbb{R}^p$ defined by $x \rightarrow (h(x), h(f(x)), \dots, h(f^p(x)))$ is 1 – 1? It is known that for a typical smooth h the mapping φ is 1 – 1 for $p > 2\dim_B(\Lambda)$ (Casdagli et al. 1991).

Denjoy Systems

This section will give some ideas indicating the principle difficulties that arise in systems with low complexity. Contrary to hyperbolic systems (each vector in the tangent space is either contracted or expanded) finer mechanisms determine the local

behavior of the scaling of balls. While in hyperbolic systems the dynamical scaling of small balls is exponential in a low-complexity system this scaling is subexponential and hence the linearization error is of the same magnitude. Up to now there is no general dimension theory for low complexity systems.

A specific example presented here is considered in (Kra and Schmeling 2002). Poincaré showed that to each orientation preserving homeomorphism of the circle $S^1 = \mathbb{R}/\mathbb{Z}$ is associated a unique real parameter $\alpha \in [0, 1)$, called the rotation number, so that the ordered orbit structure of T is the same as that of the rigid rotation R_α , where $R_\alpha(t) = (t + \alpha) \bmod 1$, provided that α is irrational. Half a century later, Denjoy (1932) constructed examples of C^1 diffeomorphisms that are not conjugate (via a homeomorphism) to rotations. This was improved later on by Herman (1979). In these examples, the minimal set of T is necessarily a Cantor set Ω .

The arithmetic properties of the rotation number have a strong effect on the properties of T . One area that has been well understood is the relation between the differentiability of T , the differentiability of the conjugation and the arithmetic properties of the rotation number. (See, for example, Herman (1979)) Without stating any precise theorem, the results differ sharply for Diophantine and for Liouville rotation numbers (definition follows). Roughly speaking the conjugating map is always regular for Diophantine rotation numbers while it might be not smooth at all for Liouville rotation numbers.

Definition 2 An irrational number α is of Diophantine class $v = v(\alpha) \in \mathbb{R}^+$ if

$$\|q\alpha\| < \frac{1}{q^\mu}$$

has infinitely many solutions in integers q for $\mu < v$ and at most finitely many for $\mu > v$ where $\|\cdot\|$ denotes the distance to the nearest integer.

In (Kra and Schmeling 2002) the effect of the rotation number on the dimension of Ω is studied. There the main result is

Theorem 6 Assume that $0 < \delta < 1$ and that $\alpha \in (0, 1)$ is of Diophantine class $v \in (0, \infty)$. Then an orientation preserving $C^{1+\delta}$ diffeomorphism of the circle with rotation number α and minimal set Ω_α^δ satisfies

$$\dim_H \Omega_\alpha^\delta \geq \frac{\delta}{v}.$$

Furthermore, these results are sharp, i.e. the standard Denjoy examples attain the minimum.

Return Times and Dimension

Recently an interesting connection between the pointwise dimensions, multifractal analysis, and recurrence behavior of trajectories was discovered (Afraimovich et al. 2000; Barreira and Saussol 2001a; Boshernitzan 1993). Roughly speaking, given an ergodic probability measure μ the return time asymptotics (as the neighborhood of the point shrinks) of μ -a.e point is determined by the pointwise dimension of μ at this point. The deeper understanding of this relation would help to get a unified approach to dimensions, exponents, entropies, recurrence times and correlation decay.

Dimension Theory of Low-Dimensional Dynamical Systems – Young's Dimension Formula

In this section a remarkable extension of Ruelle's dimension formula by Young (1982) for the dimension of a measure is discussed.

Theorem 7 Let $T: M^2 \rightarrow M^2$ be a C^2 surface diffeomorphism and let μ be an ergodic measure. Then

$$\dim_H(\mu) = h_\mu(f) \left(\frac{1}{\lambda^1} - \frac{1}{\lambda^2} \right),$$

where $\lambda^1 \leq \lambda^2$ are the two Lyapunov exponents for μ .

In (McCluskey and Manning 1983), Manning and McCluskey prove the following dimension formula for a basic set (horseshoe) of an Axiom-A surface diffeomorphism which is a set-version of Young's formula.

Theorem 8 *Let Λ be a basic set for a C^2 Axiom-A surface diffeomorphism $T : M^2 \rightarrow M^2$. Then $\dim_H(\Lambda) = s_1 + s_2$, where s_1 and s_2 satisfy*

$$P\left(-s \log \left\|DT_x|E_x^u\right\|\right) = 0$$

$$P\left(s \log \left\|DT_x|E_x^s\right\|\right) = 0.$$

where E^s and E^u are the stable and unstable directions, respectively.

Some Remarks on Dimension Theory for Low-Dimensional Versus High-Dimensional Dynamical Systems

Unlike lower dimensions (one, two, or conformal repellers), for higher-dimensional dynamical systems there are no general dimension formulas (besides the Ledrappier–Young formula), and in general dimension theory is much more difficult. This is due to several problems:

1. The geometry of the Bowen balls differs in a substantial way from round balls.
2. Number theoretic properties of some scaling rates (Pollicott and Weiss 1994; Przytycki and Urbański 1989) enter into dimension calculations in ways they do not in low dimensions (see section “Iterated Function Systems”).
3. The dimension theory of sets is often reduced to the theory of invariant measures. However, there is no invariant measure of full dimension in general and measure-theoretic considerations do not apply (McCluskey and Manning 1983).
4. The stable and unstable foliations for higher dimensional systems are typically not C^1 (Hasselblatt 1994; Schmeling 1994). Hence, to split the system into an expanding and a contracting part is far more subtle.

Dimension Theory of Higher-Dimensional Dynamical Systems

Here an example of a hyperbolic attractor in dimension 3 is considered to highlight some of the difficulties. Let Δ denote the unit disc in $\mathbb{R}^2$. Let $f : S^1 \times \Delta \rightarrow S^1 \times \Delta$ be of the form

$$T(t, x, y) = (\varphi(t), \psi^1(t, x), \psi^2(t, y)),$$

with

$$0 < \max_{S^1 \times \Delta} \frac{\partial}{\partial x} \psi^1(t, x) < \min_{S^1 \times \Delta} \frac{\partial}{\partial y} \psi^2(t, y) < \lambda < 1.$$

The limit set

$$\Lambda := \bigcap_{n \in \mathbb{N}} T^n(S^1 \times \Delta)$$

is called the attractor or the *solenoid*. It is an example of a structurally stable basic set and is one of the fundamental examples of a uniformly hyperbolic attractor.

The following result can be proved.

Theorem 9 (Bothe 1995; Hasselblatt and Schmeling 2004) *For all t , the thermodynamic pressure*

$$P\left(\dim_H \Lambda_t^s \log \left| \frac{\partial}{\partial y} \psi^2(t, y) \right| \right) = 0.$$

In particular, the stable dimension is independent of the stable section.

In this particular case the invariant axes for strong and weak contraction split the system smoothly and the difficulty is to show that the strong contraction is dominated by the weaker. In particular one has to ensure that effects as described in section “Iterated Function Systems” do not appear. In the general situation this is not the case and one lacks a similar theorem in the general situation. In particular, the unstable

foliation is not better than Hölder and does not provide a “nice” coordinate.

Hyperbolic Measures

Given $x \in M$ and a vector $v \in T_x M$ the **Lyapunov exponent** is defined as

$$\chi(x, v) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \|D_x T^n v\|.$$

provided this limit exists. For fixed x , the numbers $\chi(x, \cdot)$ attain only finitely many values. By ergodicity they are μ -a.e. constant, and the corresponding values are denoted by

$$\chi_1 \leq \cdots \leq \chi_p,$$

where $p = \dim M$. Denote by s (s stands for stable) the largest index such that $\chi_s < 0$.

Definition 3 The invariant ergodic measure μ is said to be **hyperbolic** if $\chi_i \neq 0$ for every $i = 1, \dots, p$.

The Kaplan–Yorke Conjecture

In this section a heuristic argument for the dimension of an invariant ergodic measure will be given. This argument uses a specific cover of a dynamical ball by “round” balls. These ideas are essentially developed by Kaplan and Yorke (1979) for the estimation of the dimension of an invariant set. Their estimates always provide an upper bound for the dimension. Kaplan and Yorke conjectured that in typical situations these estimates also provide a lower bound for the dimension of an attractor. Ledrappier and Young (1985a, b) showed that in case this holds for an invariant measure, then this measure has to be very special: an SBR-measure (after Sinai, Ruelle, and Bowen) (an SRB-measure is a measure that describes the asymptotic behavior for a set of initial points of positive Lebesgue measure and has absolutely continuous conditional measures on unstable manifolds).

Consider a small ball B in the phase space. The image $T^n B$ is almost an ellipsoid with axes of length

$$e^{\chi_1 n}, \dots, e^{\chi_p n}.$$

For $1 \leq i \leq s$, cover $T^n B$ by balls of radius $e^{\chi_i n}$. Then approximately

$$\frac{\exp[\chi_{i+1} n]}{\exp[\chi_i n]} \cdots \frac{\exp[\chi_p n]}{\exp[\chi_i n]}.$$

balls are needed for covering. The dimension can be estimated from above by

$$\dim_B \Lambda \leq \frac{\sum_{j>i} \chi_j}{|\chi_i|} + (p-i) := \dim_L^i \Lambda. \quad (1)$$

This is the Kaplan–Yorke formula.

General Theory

In this section the dimension theory of higher-dimensional dynamical systems is investigated. Most developed is this theory for invariant measures. There is an important connection between Lyapunov exponents and the measure theoretic entropy and dimensions of measures that will be presented here.

Let μ be an ergodic invariant measure. The Oseledec and Pesin Theory guarantee that local stable manifolds exist at μ -a.e. point. As for the Kaplan–Yorke formula the idea is to consider the contributions to the entropy and to the dimension in the directions of χ_i .

Historically the first connections between exponents and entropy were discovered by Margulis and Ruelle. They proved that

$$h_\mu(f) \leq - \sum_{i=1}^s \chi_i$$

for a C^1 diffeomorphism T . Pesin (1977) showed that this inequality is actually an equality if the measure μ is essentially the Riemannian volume on unstable manifolds. Ledrappier and Young

(1985a) showed that this is indeed a necessary condition. They also provided an exact formula:

Theorem 10 (Ledrappier–Young (1985a, b))

With $d^0 = 0$ for a C^2 diffeomorphism holds

$$h_\mu(f) = - \sum_{i=1}^s \chi_i (d^i - d^{i-1})$$

where d^i are the dimensions of the (conditional) measure on the i th unstable leaves.

The proof of this theorem is difficult and uses the theory of nonuniform hyperbolic systems (Pesin theory).

In dimension 1 and 2 the above theorem resembles Ruelle's and Young's theorems.

The reader should note that the above theorem includes also the existence of the pointwise dimension along the stable and unstable direction. Here the question arises whether this implies the existence of the pointwise dimension itself.

The Existence of the Pointwise Dimension for Hyperbolic Measure – The Eckmann–Ruelle Conjecture

In (Barreira et al. 1999), Barreira, Pesin, and Schmeling prove that every hyperbolic measure has an *almost* local product structure, i.e., the measure of a small ball can be approximated by the product of the stable conditional measure of the stable component and the unstable conditional measure of the unstable component, up to small exponential error. This was used to prove the existence of the pointwise dimension of every hyperbolic measure almost everywhere. Moreover, the pointwise dimension is the sum of the contributions from its stable and unstable part. This implies that most dimension-type characteristics of the measure (including the Hausdorff dimension, box dimension, and information dimension) coincide, and provides a rigorous mathematical justification of the concept of fractal dimension for hyperbolic measures. The existence of the pointwise dimension for hyperbolic measures had been conjectured long before by Eckmann and Ruelle.

The hypotheses of this theorem are sharp. Ledrappier and Misiurewicz (1985) constructed an example of a nonhyperbolic measure for which the pointwise dimension is not constant a.e. In (Pesin and Weiss 1997a), Pesin and Weiss present an example of a Hölder homeomorphism with Hölder constant arbitrarily close to one, where the pointwise dimension for the unique measure of maximal entropy does not exist a.e. There is also a one-dimensional example by Cutler (1990).

Endomorphisms

The previous section indicates that the dimensional properties of hyperbolic measures under invariant conditions for a diffeomorphism are understood. However, partial differential equations often generate only semi-flows and the corresponding dynamical system is noninvertible. Also, Poincaré sections sometimes introduce singularities. For such dynamical systems the theory of diffeomorphisms does not apply. However, the next theorem allows under some conditions application of this theory. It essentially rules out similar situations as considered in section “[Iterated Function Systems](#)”.

Definition 4 A system (possibly with singularities) is *almost surely invertible* if it is invertible on a set of full measure. This implies that a full measure set of points has unique forward and backward trajectories.

Theorem 11 (Schmeling–Troubetzkoy (1998))

A two-dimensional system with singularities is almost surely invertible (w.r.t. an SRB-measure) if and only if Young's formula holds.

Multifractal Analysis

A group of physicists (Halsey et al. 1986) suggested the idea of a multifractal analysis.

The Dynamical Characteristic View of Multifractal Analysis

The aim of multifractal analysis is an attempt to understand the fine structure of the level sets of the fundamental asymptotic quantities in ergodic theory (e.g., Birkhoff averages, local entropy, Lyapunov exponents). For ergodic measures these quantities are a.e. constant, however may depend on the underlying ergodic measure. Important elements of multifractal analysis entail determining the range of values these characteristics attain, an analysis of the dimension of the level sets, and an understanding of the sets where the limits do not exist. A general concept of multifractal analysis was proposed by Barreira, Schmeling and Pesin (1997b). An important field of applications of multifractal analysis is to describe sets of real numbers that have constraints on their digits or continued fraction expansion.

General Multifractal Formalism

In this section the abstract theory of multifractal analysis is described. Let X, Y be two measurable spaces and $g : X \setminus \mathcal{B} \rightarrow Y$ be any measurable function where $\mathcal{B}$ is a measurable (possibly empty) subset of X (in the standard applications $Y = \mathbb{R}$ or $Y = \mathbb{C}$). The associated **multifractal decomposition** of X is defined as

$$X = \mathcal{B} \cup \bigcup_{\alpha \in Y} K_{\alpha}^g$$

where

$$K_{\alpha}^g := \{x \in X : g(x) = \alpha\}$$

For a given set function $G : 2^X \rightarrow \mathbb{R}$ the **multifractal spectrum** is defined by

$$\mathcal{F}(\alpha) := G(K_{\alpha}^g).$$

At this point some classical and concrete examples of this general framework are considered.

The Entropy Spectrum

Let μ be an ergodic invariant measure for $T : X \rightarrow X$. If one sets

$$g^E(x) := h_{\mu}(x) \in Y = \mathbb{R}$$

and

$$G^E(Z) = h_{\text{top}}(T|Z)$$

the associated multifractal spectrum $f_{\mu}^E(\alpha)$ is called the **entropy spectrum**.

The Dimension Spectrum

This is the **classical multifractal spectrum**.

Let μ be an invariant ergodic measure on a complete separable metric space X .

Set

$$g^D(x) := d_{\mu}(x) \in Y = \mathbb{R}$$

and

$$G^D(Z) = \dim_{\text{H}} Z.$$

The associated multifractal spectrum $f_{\mu}^D(\alpha)$ is called the **dimension spectrum**.

The Lyapunov Spectrum

Let

$$g^L(x) := \chi(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \psi(T^k x) \in Y = \mathbb{R}$$

where $\psi(x) = \log |D_x T|$ and

$$G^D(Z) = \dim_{\text{H}} Z \quad \text{or} \quad G^E(Z) = h_{\text{top}}(\sigma|Z).$$

The associated multifractal spectra $f_L^D(\alpha)$ and $f_L^E(\alpha)$ are called **Lyapunov spectra**.

It was observed by H. Weiss (1999) and Barreira, Pesin and Schmeling (1997b) that for conformal repellers

$$f_L^D(\alpha) = f_{\mu_h}^D\left(\frac{\log 2}{\alpha}\right) \quad (2)$$

where the dimension spectrum on the right-hand side is with respect to the measure of maximal entropy, and

$$f_L^E(\alpha) = f_{\mu_d}^E(\dim_H \Lambda \cdot \alpha) \quad (3)$$

where the entropy spectrum on the right-hand side is with respect to the measure of maximal dimension.

The following list summarizes the state of the art for the dynamical characteristic multifractal analysis of dynamical systems. The precise statements can be found in the original papers.

- (Pesin and Weiss 1997b; Weiss 1999) For conformal repellers and Axiom-A surface diffeomorphisms, a complete multifractal analysis exists for the Lyapunov exponent.
- (Barreira et al. 1997b; Pesin and Weiss 1997b; Weiss 1999) For mixing a subshift of finite type, a complete multifractal analysis exists for the Birkhoff average for a Hölder continuous potential and for the local entropy for a Gibbs measure with Hölder continuous potential.
- (Barreira and Saussol 2001b; Pesin and Sadovskaya 2001) There is a complete multifractal analysis for hyperbolic flows.
- (Takens and Verbitski 1999) There is a generalization of the multifractal analysis on subshifts with specification and continuous potentials.
- (Barreira and Saussol 2001c; Barreira et al. 2002b) There is an analysis of “mixed” spectra like the dimension spectrum of local entropies and also an analysis of joint level sets determined by more than one (measurable) function.
- (Pollicott and Weiss 1999) For the Gauss map (and a class of nonuniformly hyperbolic maps) a complete multifractal analysis exists for the Lyapunov exponent.
- (Iommi 2005) A general approach to multifractal analysis for repellers with countably many branches is developed. It shows in contrast to finitely many branches features of non-analytic behavior.

In the first three statements the multifractal spectra are analytic concave functions that can be computed by means of the Legendre transform

of the pressure functional with respect to a suitable chosen family of potentials. In the remaining items this is no longer the case. Analyticity and convexity properties of the pressure functional are lost. However, the authors succeeded to provide a satisfactory theory in these cases.

Multifractal Analysis and Large Deviation Theory

There are deep connections between large deviation theory and multifractal analysis. The variational formula for pressure is an important tool in the analysis, and can be viewed (and proven) as a large deviation result (Ellis 1985). Some authors use large deviation theory as a tool to effect multifractal analysis.

Future Directions

The dimension theory is fast developing and of great importance in the theory of dynamical systems. In the most ideal situations (low dimensions and hyperbolicity) a generally far reaching and powerful theory has been developed. It uses ideas from statistical physics, fractal geometry, probability theory and other fields.

Unfortunately, the richness of this theory does not carry over to higher-dimensional systems. However, recent developments have shown that it is possible to obtain a general theory for the dimension of measures. Part of this theory is the development of the analytic tools of nonuniformly hyperbolic systems.

Therefore, the dimension theory of dynamical systems is far from complete. In particular, it is usually difficult to apply the general theory to concrete examples, for instance if one really wants to *compute* the dimension. The general theory does not provide a way to compute the dimension but gives rather connections to other characteristics. Moreover, in the presence of neutral directions (zero Lyapunov exponents) one encounters all the difficulties arising in low-complexity systems.

Another important open problem is to understand the dimension theory of invariant sets in higher-dimensional spaces. One way would be to

relate the dimension of sets to the dimension of measures. Such a connection is not clear. The reason is that most systems do not exhibit a measure whose dimension coincides with the dimension of its support (invariant set). But there are some reasons to conjecture that any compact invariant set of an expanding map in any dimension carries a measure of maximal dimension (see (Gatzouras and Peres 1996; Kenyon and Peres 1996)). If this conjecture is true one obtains an invariant measure whose unstable dimension coincides with the unstable dimension of the invariant set. There is also a measure of maximal stable dimension. Combining these two measures one could establish an analogous theory for invariant sets as for invariant measures.

Last but not least one has to mention the impact of the dimension theory of dynamical systems on other fields. This new point of view makes in many cases the posed problems more tractable. This is illustrated in examples from number theory, geometric limit constructions and others. The applications of the dimension theory of dynamical systems to other questions seem to be unlimited.

Bibliography

Primary Literature

- Afraimovich VS, Chernov NI, Sataev EA (1995) Statistical properties of 2D generalized hyperbolic attractors. *Chaos* 5:238–252
- Afraimovich VS, Schmeling J, Ugalde J, Urias J (2000) Spectra of dimension for Poincaré recurrences. *Discret Contin Dyn Syst* 6(4):901–914
- Aihua F, Yunping J, Jun W (2005) Asymptotic Hausdorff dimensions of Cantor sets associated with an asymptotically non-hyperbolic family. *Ergodic Theory Dynam Syst* 25(6):1799–1808
- Alexander J, Yorke J (1984) Fat Baker's transformations. *Ergodic Theory Dynam Syst* 4:1–23
- Ambroladze A, Schmeling J (2004) Lyapunov exponents are not stable with respect to arithmetic subsequences. In: *Fractal geometry and stochastics III*. Progr Probab, vol 57. Birkhäuser, Basel, pp 109–116
- Artin E (1965) Ein mechanisches System mit quasi-ergodischen Bahnen, collected papers. Addison Wesley, pp 499–501
- Barreira L (1995) Cantor sets with complicated geometry and modeled by general symbolic dynamics. *Random Comput Dyn* 3:213–239
- Barreira L (1996a) A non-additive thermodynamic formalism and applications to dimension theory of hyperbolic dynamical systems. *Ergodic Theory Dynam Syst* 16: 871–927
- Barreira L (1996b) A non-additive thermodynamic formalism and dimension theory of hyperbolic dynamical systems. *Math Res Lett* 3:499–509
- Barreira L. Variational properties of multifractal spectra. IST preprint
- Barreira L, Saussol B (2001a) Hausdorff dimension of measure via Poincaré recurrence. *Commun Math Phys* 219(2):443–463
- Barreira L, Saussol B (2001b) Multifractal analysis of hyperbolic flows. *Commun Math Phys* 219(2):443–463
- Barreira L, Saussol B (2001c) Variational principles and mixed multifractal spectra. *Trans Am Math Soc* 353(10):3919–3944. (electronic)
- Barreira L, Schmeling J (2000) Sets of “non-typical” points have full topological entropy and full Hausdorff dimension. *Israel J Math* 116:29–70
- Barreira L, Pesin Y, Schmeling J (1997a) Multifractal spectra and multifractal rigidity for horseshoes. *J Dyn Contr Syst* 3:33–49
- Barreira L, Pesin Y, Schmeling J (1997b) On a general concept of multifractal rigidity: multifractal spectra for dimensions, entropies, and Lyapunov exponents. *Multifractal rigidity*. *Chaos* 7:27–38
- Barreira L, Pesin Y, Schmeling J (1999) Dimension and product structure of hyperbolic measures. *Ann Math* 149:755–783
- Barreira L, Saussol B, Schmeling J (2002a) Distribution of frequencies of digits via multifractal analysis. *J Number Theory* 97(2):410–438
- Barreira L, Saussol B, Schmeling J (2002b) Higher-dimensional multifractal analysis. *J Math Pures Appl* (9) 81(1):67–91
- Belykh VP (1982) Models of discrete systems of phase synchronization. In: Shakhildyan VV, Belyushina LN (eds) *Systems of phase synchronization*. Radio i Svyaz, Moscow, pp 61–176
- Billingsley P (1978) *Ergodic theory and information*. Krieger
- Blinchevskaya M, Ilyashenko Y (1999) Estimate for the entropy dimension of the maximal attractor for k -contracting systems in an infinite-dimensional space. *Russ J Math Phys* 6(1):20–26
- Boshernitzan M (1993) Quantitative recurrence results. *Invent Math* 113:617–631
- Bothe H-G (1995) The Hausdorff dimension of certain solenoids. *Ergodic Theory Dynam Syst* 15:449–474
- Bousch T (2000) Le poisson n'a pas d'arêtes. *Ann IH Poincaré (Prob-Stat)* 36(4):489–508
- Bowen R (1973) Topological entropy for noncompact sets. *Trans Am Math Soc* 184:125–136
- Bowen R (1979) Hausdorff dimension of quasi-circles. *Publ Math IHES* 50:11–25
- Bylov D, Vinograd R, Grobman D, Nemyckii V (1966) Theory of Lyapunov exponents and its application to

- problems of stability. Izdat "Nauka", Moscow (in Russian)
- Casdagli M, Sauer T, Yorke J (1991) Embedology. *J Stat Phys* 65:589–616
- Ciliberto S, Eckmann JP, Kamphorst S, Ruelle D (1971) Liapunov exponents from times. *Phys Rev A* 34
- Collet P, Lebowitz JL, Porzio A (1987) The dimension spectrum of some dynamical systems. *J Stat Phys* 47: 609–644
- Constantin P, Foias C (1988) Navier-Stokes equations. Chicago University Press
- Cruchfield J, Farmer D, Packard N, Shaw R (1980) Geometry from a time series. *Phys Rev Lett* 45:712–724
- Cutler C (1990) Connecting ergodicity and dimension in dynamical systems. *Ergodic Theory Dynam Syst* 10: 451–462
- Denjoy A (1932) Sur les courbes définies par les équations différentielles à la surface du tore. *J Math Pures Appl* 2: 333–375
- Ding M, Grebogi C, Ott E, Yorke J (1993) Estimating correlation dimension from a chaotic times series: when does the plateau onset occur? *Phys D* 69:404–424
- Dodson M, Rynne B, Vickers J (1990) Diophantine approximation and a lower bound for Hausdorff dimension. *Mathematika* 37:59–73
- Douady A, Oesterle J (1980) Dimension de Hausdorff Des Attracteurs. *CRAS* 290:1135–1138
- Eggleston HG (1952) Sets of fractional dimension which occur in some problems of number theory. *Proc Lond Math Soc* 54:42–93
- Ellis R (1984) Large deviations for a general class of random vectors. *Ann Prob* 12:1–12
- Ellis R (1985) Entropy, large deviations, and statistical mechanics. Springer
- Falconer K (1990) Fractal geometry, mathematical foundations and applications. Cambridge University Press, Cambridge
- Fan AH, Feng DJ, Wu J (2001) Recurrence, dimension and entropy. *J Lond Math Soc* 64(2):229–244
- Frederickson P, Kaplan J, Yorke E, Yorke J (1983) The Liapunov dimension of strange attractors. *J Differ Equ* 49:185–207
- Frostman O (1935) Potential d'équilibre Et Capacité des Ensembles Avec Quelques Applications à la Théorie des Fonctions. *Meddel Lunds Univ Math Semin* 3: 1–118
- Furstenberg H (1967) Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math Syst Theory* 1:1–49
- Furstenberg H (1970) Intersections of Cantor sets and transversality of semigroups I. In: Problems in analysis. Sympos Salomon Bochner. Princeton University Press, pp 41–59
- Gatzouras D, Peres Y (1996) The variational principle for Hausdorff dimension: a survey. In: Pollicott M et al (eds) Ergodic theory of Z^d actions. Proc of the Warwick symposium. Lond Math Soc Lect Note Ser vol 228. Cambridge University Press, pp 113–125
- Grassberger P, Procaccia I, Hentschel H (1983) On the characterization of chaotic motions, Lect notes. *Physics* 179:212–221
- Halsey T, Jensen M, Kadanoff L, Procaccia I, Shraiman B (1986) Fractal measures and their singularities: the characterization of strange sets. *Phys Rev A* 33(N2):1141–1151
- Hasselblatt B (1994) Regularity of the Anosov splitting and of horospheric foliations. *Ergodic Theory Dynam Syst* 14(4):645–666
- Hasselblatt B, Schmeling J (2004) Dimension product structure of hyperbolic sets. In: Modern dynamical systems and applications. Cambridge University Press, Cambridge, pp 331–345
- Henley D (1992) Continued fraction cantor sets, Hausdorff dimension, and functional analysis. *J Number Theory* 40:336–358
- Hentschel HGE, Procaccia I (1983) The infinite number of generalized dimensions of fractals and strange attractors. *Physica* 8D:435–444
- Herman MR (1979) Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations. *Publ l'Inst Math Hautes Études Sci* 49:5–234
- Hunt B (1996) Maximal local Lyapunov dimension bounds the box dimension of chaotic attractors. *Nonlinearity* 9: 845–852
- Iommi G (2005) Multifractal analysis for countable Markov shifts. *Ergodic Theory Dynam Syst* 25(6): 1881–1907
- Jarnik V (1931) Über die simultanen diophantischen Approximationen. *Math Zeitschr* 33:505–543
- Jenkinson O (2001) Rotation, entropy, and equilibrium states. *Trans Am Math Soc* 353:3713–3739
- Kalinin B, Sadovskaya V (2002) On pointwise dimension of non-hyperbolic measures. *Ergodic Theory Dynam Syst* 22(6):1783–1801
- Kaplan JL, Yorke JA (1979) Functional differential equations and approximation of fixed points. Lecture notes. In: Mathematics, vol 730. Springer, Berlin, pp 204–227
- Katznelson Y, Weiss B (1982) A simple proof of some ergodic theorems. *Isr J Math* 42:291–296
- Kenyon R, Peres Y (1996) Measure of full dimension on affine invariant sets. *Ergodic Theory Dynam Syst* 16: 307–323
- Kesseböhmer M (1999) Multifrakale und Asymptotiken grosser Deviationen. Thesis U Göttingen, Göttingen
- Kingman JFC (1968) The ergodic theory of subadditive stochastic processes. *J R Stat Soc B30*:499–510
- Kleinbock D, Margulis G (1998) Flows on homogeneous spaces and Diophantine approximation on manifold. *Ann Math* 148:339–360
- Kra B, Schmeling J (2002) Diophantine classes, dimension and Denjoy maps. *Acta Arith* 105(4):323–340
- Ledrappier F (1981) Some relations between dimension and Lyapounov exponents. *Commun Math Phys* 81: 229–238
- Ledrappier F (1986) Dimension of invariant measures. Proceedings of the conference on ergodic theory and

- related topics II (Georgenthal, 1986). Teubner-texte Math 94:137–173
- Ledrappier F, Misiurewicz M (1985) Dimension of invariant measures for maps with exponent zero. *Ergodic Theory Dynam Syst* 5:595–610
- Ledrappier F, Strelcyn JM (1982) A proof of the estimate from below in Pesin's entropy formula. *Ergodic Theory Dynam Syst* 2:203–219
- Ledrappier F, Young LS (1985a) The metric entropy of diffeomorphisms. I Characterization of measures satisfying Pesin's entropy formula. *Ann Math* 122:509–539
- Ledrappier F, Young LS (1985b) The metric entropy of diffeomorphisms, II. Relations between entropy, exponents and dimension. *Ann Math* 122:540–574
- Lopes A (1989) The dimension spectrum of the maximal measure. *SIAM J Math Anal* 20:1243–1254
- Mănă R (1981) On the dimension of compact invariant sets for certain nonlinear maps. *Lecture notes in mathematics*, vol 898. Springer
- Mănă R (1990) The Hausdorff dimension of horseshoes of surfaces. *Bol Soc Bras Math* 20:1–24
- Mauldin RD, Urbański M (1996) Dimensions and measures in infinite iterated function systems. *Proc Lond Math Soc* 73(1):105–154
- Mauldin RD, Urbański M (2000) Parabolic iterated function systems. *Ergodic Theory Dynam Syst* 20(5):1423–1447
- Mauldin RD, Urbański M (2002) Fractal measures for parabolic IFS. *Adv Math* 168(2):225–253
- Mauldin DR, Urbański M (2003) Graph directed Markov systems. In: *Geometry and dynamics of limit sets*. Cambridge tracts in mathematics, vol 148. Cambridge University Press, Cambridge
- McCluskey H, Manning A (1983) Hausdorff dimension For horseshoes. *Ergodic Theory Dynam Syst* 3:251–260
- Moran P (1946) Additive functions of intervals and Hausdorff dimension. *Proc Cambr Philos Soc* 42:15–23
- Moreira C, Yoccoz J (2001) Stable intersections of cantor sets with large Hausdorff dimension. *Ann Math* 154(1):45–96
- Neunhäuserer J (1999) An analysis of dimensional theoretical properties of some affine dynamical systems. Thesis, Free University Berlin, Berlin
- Oseledets V (1968) A multiplicative ergodic theorem. Liapunov characteristic numbers for dynamical systems. *Trans Moscow Math Soc* 19:197–221
- Ott E, Sauer T, Yorke J (1994) Part I Background. In: *Coping with chaos*. Wiley Ser Nonlinear Sci. Wiley, New York, pp 1–62
- Palis J, Takens F (1987) Hyperbolicity and the creation of homoclinic orbits. *Ann Math* 125:337–374
- Palis J, Takens F (1993) Hyperbolicity and sensitive chaotic dynamics at homoclinic bifurcations. Cambridge University Press, Cambridge
- Palis J, Takens F (1994) Homoclinic tangencies for hyperbolic sets of large Hausdorff dimension. *Acta Math* 172:91–136
- Palis J, Viana M (1988) On the continuity of Hausdorff dimension and limit capacity for horseshoes. *Lecture notes in math*, vol 1331. Springer
- Pesin Y (1977) Characteristic exponents and smooth ergodic theory. *Russ Math Surv* 32(4):55–114
- Pesin Y, Pitskel B (1984) Topological pressure and the variational principle for noncompact sets. *Funct Anal Appl* 18:307–318
- Pesin Y (1992) Dynamical systems with generalized hyperbolic attractors: hyperbolic, ergodic and topological properties. *Ergodic Theory Dynam Syst* 12:123–151
- Pesin Y (1993) On rigorous mathematical definition of correlation dimension and generalized spectrum for dimensions. *J Stat Phys* 71(3–4):529–547
- Pesin Y (1997a) Dimension theory in dynamical systems: rigorous results and applications. Cambridge University Press, Cambridge
- Pesin Y (1997b) Dimension theory in dynamical systems: contemporary views and applications. In: *Chicago lectures in mathematics*. Chicago University Press, Chicago
- Pesin Y, Sadovskaya V (2001) Multifractal analysis of conformal axiom A flows. *Commun Math Phys* 216(2):277–312
- Pesin Y, Tempelman A (1995) Correlation dimension of measures invariant under group actions. *Random Comput Dyn* 3(3):137–156
- Pesin Y, Weiss H (1994) On the dimension of deterministic and random cantor-like sets. *Math Res Lett* 1:519–529
- Pesin Y, Weiss H (1996) On the dimension of deterministic and random cantor-like sets, symbolic dynamics, and the Eckmann-Ruelle conjecture. *Commun Math Phys* 182:105–153
- Pesin Y, Weiss H (1997a) A multifractal analysis of equilibrium measures for conformal expanding maps and Moran-like geometric constructions. *J Stat Phys* 86:233–275
- Pesin Y, Weiss H (1997b) The multifractal analysis of Gibbs measures: motivation. *Mathematical foundation and examples*. *Chaos* 7:89–106
- Petersen K (1983) *Ergodic theory*. Cambridge studies in advanced mathematics 2. Cambridge University Press, Cambridge
- Pollicott M, Weiss H (1994) The dimensions of some self affine limit sets in the plane and hyperbolic sets. *J Stat Phys* 77:841–866
- Pollicott M, Weiss H (1999) Multifractal analysis for the continued fraction and Manneville-Pomeau transformations and applications to Diophantine approximation. *Commun Math Phys* 207(1):145–171
- Przytycki F, Urbański M (1989) On Hausdorff dimension of some fractal sets. *Stud Math* 93:155–167
- Ruelle D (1978) *Thermodynamic formalism*. Addison-Wesley
- Ruelle D (1982) Repellers for real analytic maps. *Ergodic Theory Dynam Syst* 2:99–107

- Schmeling J (1994) Hölder continuity of the Holonomy maps for hyperbolic basic sets II. *Math Nachr* 170: 211–225
- Schmeling J (1997) Symbolic dynamics for Beta-shifts and self-normal numbers. *Ergodic Theory Dynam Syst* 17: 675–694
- Schmeling J (1998) A dimension formula for endomorphisms – the Belykh family. *Ergodic Theory Dynam Syst* 18:1283–1309
- Schmeling J (1999) On the completeness of multifractal spectra. *Ergodic Theory Dynam Syst* 19:1–22
- Schmeling J (2001) Entropy Preservation under Markov codings. *J Stat Phys* 104(3–4):799–815
- Schmeling J, Troubetzkoy S (1998) Dimension and invertibility of hyperbolic endomorphisms with singularities. *Ergodic Theory Dynam Syst* 18: 1257–1282
- Schmeling J, Weiss H (2001) Dimension theory and dynamics. *AMS Proc Symposia Pure Math Ser* 69: 429–488
- Series C (1985) The modular surface and continued fractions. *J Lond Math Soc* 31:69–80
- Simon K (1997) The Hausdorff dimension of the general Smale–Williams solenoid with different contraction coefficients. *Proc Am Math Soc* 125:1221–1228
- Simpelaere D (1994) Dimension spectrum of axiom-A diffeomorphisms, II. Gibbs measures. *J Stat Phys* 76: 1359–1375
- Solomyak B (1995) On the random series $\sum \pm \lambda^n$ (an Erdős problem). *Ann Math* 142(3):611–625
- Solomyak B (2004) Notes on Bernoulli convolutions. In: *Fractal geometry and applications: a jubilee of Benoit Mandelbrot*. *Proc Sympos Pure Math*, vol 72, Part 1. American Mathematical Society, Providence, pp 207–230
- Stratmann B (1995) Fractal dimensions for Jarnik limit sets of geometrically finite Kleinian groups; the semi-classical approach. *Ark Mat* 33:385–403
- Takens F (1981) Detecting strange attractors in turbulence. *Lecture notes in mathematics*, vol 898. Springer
- Takens F, Verbitski E (1999) Multifractal analysis of local entropies for expansive homeomorphisms with specification. *Commun Math Phys* 203:593–612
- Walters P (1982) *Introduction to ergodic theory*. Springer
- Weiss H (1992) Some variational formulas for Hausdorff dimension, topological entropy, and SRB entropy for hyperbolic dynamical system. *J Stat Phys* 69:879–886
- Weiss H (1999) The Lyapunov spectrum of equilibrium measures for conformal expanding maps and axiom-A surface diffeomorphisms. *J Stat Phys* 95(3–4):615–632
- Young LS (1981) Capacity of attractors. *Ergodic Theory Dynam Syst* 1:381–388
- Young LS (1982) Dimension, entropy, and Lyapunov exponents. *Ergodic Theory Dynam Syst* 2:109–124

Books and Reviews

- Bowen R (1975) *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*. *Lecture notes in mathematics*, vol 470. Springer
- Eckmann JP, Ruelle D (1985) Ergodic theory of chaos and strange attractors. *Rev Mod Phys* 57:617–656
- Federer H (1969) *Geometric measure theory*. Springer
- Hasselblatt B, Katok A (2002) *Handbook of dynamical systems*, vol 1, Survey 1. Principal structures. Elsevier
- Katok A (1980) Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst Hautes Études Sci Publ Math* 51:137–173
- Katok A, Hasselblatt B (1995) *Introduction to the modern theory of dynamical systems*. Cambridge University Press, Cambridge
- Keller G (1998) *Equilibrium states in ergodic theory*. In: *London Mathematical Society student texts* 42. Cambridge University Press, Cambridge
- Mañé R (1987) Ergodic theory and differentiable dynamics. In: *Ergebnisse der Mathematik und ihrer Grenzgebiete* 3, vol 8. Springer
- Mario R, Urbański M (2005) Regularity properties of Hausdorff dimension in infinite conformal iterated function systems. *Ergodic Theory Dynam Syst* 25(6): 1961–1983
- Mattila P (1995) Geometry of sets and measures in Euclidean spaces. In: *Fractals and rectifiability*. Cambridge University Press, Cambridge
- Pugh C, Shub M (1989) Ergodic attractors. *Trans Am Math Soc* 312(1):1–54
- Takens F (1988) Limit capacity and Hausdorff dimension of dynamically defined Cantor sets. *Lecture notes in math*, vol 1331. Springer

Index

A

Abelian variety, 589
Absolute Cesàro boundedness, 480
Absolute continuity, 322
Absorption theorem, 501
Abstract Cauchy problem, 481
Adic transformations, 26
Adjacency matrix, 434
AF-equivalence relation, 498, 499, 501–504
Affable equivalence relation, 494, 502–504
Ageev, Danilenko, 121
Akcoglu–Krengel theorem, 91
Akcoglu’s theorem, 476
Alexeyev’s theorem, 115–116
 σ -algebra, 238
Algebraic dynamical systems, 590
Algebraic K-theory, 451, 622
Algorithmic complexity, 650
Almost conjugacy, 427, 441, 447
Almost finite actions, 517
Alpern’s lemma, 237
Ambrose–Kakutani theorem, 239
Amenable group, 214
Amenable radical, 521
Analytic number theory, 578
Analytic set, 529, 570
Anosov diffeomorphism, 16, 18, 319, 384, 541, 542, 617
Anosov flow, 16, 18, 586, 618
Anosov systems, 16–17
Aperiodicity, 293, 303, 304
Area-preserving smooth flow, 333
Arithmetic function, 293
Arithmetic progression, 582–584, 588, 593
Arnold Cat Map, 17
Arnold flows, 335, 337
 rescalings of, 361
 special representation of, 342
Asymmetric singularities, 343
Asymptotically orthogonal powers (AOP), 165, 304
Asymptotic equipartition property, 375
Attractor, 6
 reconstruction, 671
Auslander–Ellis theorem, 398

Automorphism, 81, 294, 427, 440
 group, 440
Average-distropy-number, 176
Axiom A diffeomorphisms, 384
Axiom A flow, 586
Axiom A system, 17–18

B

Baire space, 545
Baker’s theorem, 577, 590, 592
Banach lattice, 464
Banach principle, 84
Banach spaces, 642
Baumslag–Solitar group, 493
Benchmarking, 529, 534
Benedicks and Carleson analysis, 656
Bergelson’s question, 278
Bernoulli factors, 262
Bernoulli flow, 211
Bernoulli measure, 9
Bernoulli processes, 177
Bernoulli property, 51, 207
Bernoulli shift, 7, 9, 37, 39, 40, 45, 51, 56, 177, 202, 538–539, 643
 explicit isomorphisms, 205
 isomorphic transformation to, 208
 isomorphism of, 205–207
 Kolmogorov–Sinai entropy, 205
 one sided, 203
 Ornstein’s isomorphism theorem, 206–207
Berry–Esseen inequality, 649
Besicovitch sequences, 478
Bi-Lipschitz equivalence, 558
Billiard flow, 12
Billiard map, 12, 13
Billiard systems, 12–13
Binomial coefficients, 172–173
Birkhoff ergodic theorem, 64, 534
Birkhoff pointwise ergodic theorem, 44, 638
Birkhoff recurrence theorem, 392–393
Birkhoff’s theorem, 82
Birkhoff theorem, 639

- Block multiplication, 8
 - Blokhin flows, 337
 - Blum-Hanson (BH) property, 470
 - Bochner-Herglotz theorem, 111
 - Bohr recurrence, 73, 74
 - Boltzmann's ergodic hypothesis, 43
 - Boltzmann's hypothesis, 327
 - Borel classes, 547–548
 - Borel flow, 333
 - Borel hierarchy, 569–571
 - Borel measure, 19
 - Borel probability, 638
 - measure, 3, 8, 20, 374
 - Borel reducibility, 543
 - Borel set, 11
 - Borel's theorem, 9, 581
 - G -boundary, 520, 522
 - Boundary map, 520
 - Bounded gaps, 64, 69, 71
 - Bounded prime volume (BPV), 305
 - Bowen ball, 668
 - Bowen–Dinaburg formulae, 650
 - Bratteli diagram, 494, 499–502
 - Bratteli-Elliott-Krieger theorem, 501
 - Bratteli-Vershik map, 240
 - Bratteli-Vershik system, 500
 - Bratteli-Vershik transformations, 498, 500
 - Brownian motion, 220
 - Bufetov functionals, 351
 - Bunimovich billiards, 13
 - Butterfly effect, 334
- C**
- Cameron-Martin theorem, 243
 - Canonical anticommutation relations (CAR), 284
 - Cantor minimal systems, 494, 495, 497, 503, 504, 506, 508, 509, 517
 - Cantor set, 18, 21
 - Centralizer, 217
 - Central Limit Theorem, 647–650
 - Cesàro averaging, 474
 - Cesàro bounded operator, 461–463, 480
 - Chacón–Ornstein theorem, 90
 - Chacon's transformation, 158
 - Chacon system, 8
 - Chacon transformations, 38
 - Chain recurrent set, 641, 653
 - Chaos, 634
 - Chouquet simplex, 558
 - Chowla conjecture, 165, 296–298
 - Classical multifractal spectrum, 676
 - Co-analytic set, 529, 546, 570–572
 - Coarse topology, 178
 - Coded systems, 9
 - Cohomological equation, 338
 - Coin-flipping measure, 552
 - Collet–Eckmann unimodal maps, 649
 - Column, 38
 - Combinatorial freedom, 637, 652
 - Combinatorial number theory, 41–42, 391
 - Compact group rotation, 266
 - μ -compatible metric, 63, 64
 - Completely deterministic point, 293
 - Completely-positive entropy, 51, 181
 - Complete order isomorphism, 521, 522
 - Complex dynamics, 22
 - Complexity, 375
 - Computability theory, 543
 - W -condition, 352
 - Conditional distropy, 175
 - Conditional information function, 175
 - Conditional probability, 175
 - Cone method, 327
 - Conjugacy, 5, 451, 507, 508
 - problem, 437–440
 - Connes-Feldman-Weiss theorem, 492
 - Conservative system, 63
 - Conservativity, 234–236
 - Consistent temporal modeling, 42
 - Constrained systems, 435
 - Context-free systems, 9
 - Continued fraction digits, 580
 - Continued fraction map, 10
 - Continuity properties of the topological dynamics, 654
 - Continuous action, 495
 - Continuous orbit equivalence, 493, 494, 506–510
 - Continuous-time flows, 431
 - Contracting subspace, 14
 - Conze–Lesigne factor, 99
 - Countable amenable group, 512, 519
 - Countable equivalence relations, 551
 - Countable generator, 238
 - Countable Lebesgue spectrum, 50, 358
 - Counting measure, 217
 - Critical points, 644, 645
 - C_0 -semigroup, 480
 - Cuntz comparison for open sets, 514, 515
 - Curtis-Hedlund-Lyndon theorem, 433
 - Curvature, 577, 585–587
 - Cylinder set, 6, 37
- D**
- Decay of correlations, 53
 - Deep neural networks (DNNs), 42, 43
 - Deformation rigidity, 623
 - Denjoy systems, 504, 671–672
 - Derived from Anosov (DA) attractor, 320
 - Descriptive complexity, 542
 - continuous reductions, 544
 - ill-founded tree, reduction of, 545–546
 - pre-orderings, 544
 - reduction of sets, 543
 - reductions of equivalence relations, 543–544
 - Descriptive set theory, 569

- Devaney chaos, 636
 - C^1 diffeomorphism, 653
 - Differentiable dynamics, 382
 - Axiom A diffeomorphisms, Anosov diffeomorphisms, Sinai-Ruelle-Bowen measures, 383–385
 - Hausdorff dimension of conformal repellers, 385
 - interval maps, indifferent fixed point, 383
 - Markov maps of the interval, 382–383
 - Differentiable rigidity, 620–623
 - Dimension group, 427, 440
 - Dimension representation, 440
 - Diophantine analysis, 578, 579
 - mixing and additive relations, 590–592
 - orbit growth and convergence, 589–590
 - Diophantine approximation, 577, 580, 597, 599, 601, 603, 609
 - Diophantine exponent, 600, 604
 - Diophantine-like conditions, 353
 - Diophantine problem, 577, 591
 - Directional recurrence, 279
 - Dirichlet's theorem, 599
 - Disjointness, 150, 153
 - from mixing systems, 359
 - of rescalings, 360–361, 363
 - theory, 295
 - Dispersing billiards, 13
 - Dissipative partially hyperbolic diffeomorphisms, 325
 - Distortion estimates, 320
 - Distribution entropy, 171–173
 - Doebelin–Fortet, 644
 - Doubling map, 37, 63, 64, 69, 71
 - Dyadic (von Neumann-Kakutani) adding machine, 6
 - Dynamical systems, 8, 334, 532
 - complexity in structure theory, 546–550
 - complexity of classification theory, 550–556
 - descriptive complexity (*see* Descriptive complexity)
 - Hamilton dynamics, 534–535
 - measure structure, 537–540
 - presentations, 540
 - rotations of unit circle, 535–536
 - smooth dynamics, 540–542
 - symbolic shifts, 536–537
 - Dynamical zeta function, 577, 587, 589
- E**
- Eckmann–Ruelle conjecture, 675
 - Effros–Handelman–Shen Theorem, 500, 502
 - Elliott invariant, 505, 515, 516, 518
 - Ellis–Namakura theorem, 398
 - Embedding, 427, 433
 - theorem, 452
 - Endomorphism, 81, 675
 - Energy surface, 36
 - Ensemble averages, 638
 - Entropy, 51, 92, 170–171, 176, 269–272, 293, 435
 - Bernoulli processes, 177
 - determinism and zero-entropy, 180–181
 - distribution, 171–173
 - flow (R-action), 192–193
 - function, 375
 - as function of measure, 375
 - generators, 176
 - information function, 174–175
 - of invariant measures, 374–375
 - non-amenable groups, 196–198
 - Ornstein theory, 182
 - Pinsker-field and K-automorphisms, 181–182
 - production, 371, 386
 - relative entropy, 380–381
 - Shannon's noisy channel theorem, 173–174
 - skew product and random walk in random scenery, 182–184
 - spectrum, 676
 - time reversibility, 176–177
 - topological, 184–192
 - transformation, 177–180
 - Z^d -actions, 193–195
 - Entropy-expanding maps, 653, 657
 - Entropy theory
 - definition, 402
 - entropy n -tuples, 404–406
 - Kolmogorov system, 403–404
 - maximal pattern entropy, 408–409
 - null systems, 408
 - sequence entropy, 406–408
 - topological vs measurable, 402–403
 - tuples, 409
 - weak horseshoe, 406
 - Equidistribution, 577, 578, 581, 582
 - Equilibrium distribution, 372
 - Equilibrium measures, 7
 - Equilibrium states, 376–378
 - nonuniqueness of, 378
 - Equivalence relation
 - complexity of classification theory, 550–556
 - countable, 551
 - and Friedman-Stanley jump operator, 553
 - on homeomorphisms, 535
 - measure conjugacy, 533
 - natural, 535
 - Polish Group action, 555
 - reductions of, 543–544
 - S_∞ -actions, 553–555
 - topological conjugacy, 533
 - turbulence, 555
 - with uncountable classes, 551–553
 - Ergodic average, 579, 581, 582
 - Ergodic decomposition, 45, 46, 48, 69, 237–238, 394
 - theorem, 537
 - Ergodic flow, 333
 - Ergodic hypothesis, 36, 578, 579
 - Ergodicity, 36, 43–46, 394
 - of Gaussian cocycles, 280
 - of powers, 279
 - Ergodic measure preserving systems, 549, 550
 - Ergodic properties, 222
 - Ergodic Ramsey theory, 71, 98

- Ergodic Ramsey theory and recurrence, 582–583
 - polynomialization and IP -sets, 584
 - sets of primes, 584
 - topology and coloring theorems, 583–584
- Ergodic theorems, 532, 579, 582
 - definition, 79
 - local, 86
 - mean, 579
 - for measure-preserving maps, 81–85
 - for non-amenable groups, 101
 - pointwise, 88–90, 579
 - subadditive and multiplicative, 90–92
- Ergodic theory, 1, 2, 299, 446–447, 597, 598, 667
 - adding machines, 6
 - Adic transformations, 26
 - Anosov systems, 16–17
 - Axiom A system, 17–18
 - Borel probability measure, 401
 - chaos and complexity theory, 400–401
 - characteristic factors, 414–416
 - complex dynamics, 22
 - continued fraction map, 10
 - cutting and stacking, 25–26
 - disjointness, 399
 - entropy (*see* Entropy)
 - factors, 23
 - Farey map, 10
 - f -expansions, 10–11
 - full shifts and shifts of finite type, 6–8
 - Gaussian systems, 11
 - geodesic flow on manifold of negative curvature, 15
 - Hamiltonian systems, 11–13
 - history, 390
 - horocycle flow, 15–16
 - induced transformations, 24–25
 - infinite ergodic theory, 22–23
 - interval exchange maps, 6
 - inverse limits, 27
 - joinings, 28
 - Markov partitions and coding, 16
 - multiple ergodic averages, 412–414
 - natural extension, 27–28
 - non-uniformly hyperbolic systems, 19
 - partially hyperbolic dynamical systems, 18–19
 - physically relevant measures and strange attractors, 19–22
 - problems, 391
 - products, 23
 - recurrence (*see* Recurrence)
 - rigid rotation of compact group, 5–6
 - Rokhlin's lemma, 26–27
 - β -shifts, 11
 - skew products, 24
 - smooth expanding interval maps, 9–10
 - smooth uniformly hyperbolic diffeomorphisms and flows, 14–18
 - structure theorems, 411–412
 - subshifts, 8–9
 - suspension flows, 25
 - Ergodic transformation, 2, 540
 - Étale equivalence, 493, 495–499, 501, 502, 504
 - Euclidean coordinates, 12
 - Euclidean space, 14
 - Even shift, 431
 - Expanding maps, 319
 - Expanding subspace, 14
- F**
 - Factor map, 183, 197, 427, 433, 529, 537
 - Farey map, 10
 - Fatou set, 22
 - Filtering problems, 160
 - Finitary isomorphisms, 210
 - Finite equivalence, 427, 441, 442
 - Finite lattice, 373
 - Finite partition, 16
 - Finite set, 6
 - Finite-state coding theorem, 445
 - Finite statistical description, 639–641, 644, 646, 647
 - First digits, 581
 - First-return map, 24
 - Fock space, 221
 - Foiaş–Strătilă property, 267
 - Foiaş–Stratila theorem, 126–128, 229
 - Follower sets, 9
 - Følner sequence, 93, 214
 - Fractal geometry, 665, 669
 - Free action, 493, 496, 517–519
 - Friedman–Stanley jump operator, 553
 - Full shift, 427, 431, 447
 - Full tent map, 635, 636, 638, 639, 642, 650, 652
 - Furstenberg boundary, 493, 520, 521
 - Furstenberg correspondence principle, 584
 - Furstenberg disjointness, 308
 - Furstenberg ergodic averages, 71
 - Furstenberg systems, 294, 303, 307
 - Furstenberg theorem, 280
 - Furstenberg–Zimmer's structure theorem, 412
- G**
 - GAGs, 228, 229
 - Gaussian and Poisson systems, 226
 - Gaussian centralizer, 225
 - Gaussian-compact factor, 226
 - Gaussian distribution, 219
 - Gaussian dynamical system, 221, 592
 - Gaussian factor, 224
 - Gaussian-Kronecker automorphism, 124
 - Gaussian-Kronecker system, 228
 - Gaussian process, 1, 217
 - Gaussian random variable, 11
 - Gaussian space, 217
 - Gaussian stationary processes, 219
 - Gaussian systems, 11
 - Gaussian transformations, 243, 265
 - Gauss–Kuz'min law, 580
 - Gauss map, 581

Gauss measure, 10, 581
 Gelfand-Naimark theorem, 522
 General averaging methods, 474–476
 Generalized Gaussian systems, 220
 Generalized Krengel class, 260
 Generalized Morse systems, 8
 General nonsingular Bernoulli shifts, 260
 Generic point, 294
 Geodesic(s), 585–587
 flow, 578, 580, 585–587
 Geometric interpretation, 220
 Gibbs distribution, 372
 Gibbs–Markov map, 650
 Gibbs property, 372
 class of regular potentials, 379
 Perron-Frobenius theorem, 379–380
 relative entropy, 380–381
 subshifts of finite type, 378–379
 Gibbs states, 381
 Glimm-Effros dichotomy, 551–553
 Glimm-Effros theorem, 238–239
 Global-global mixing, 257
 Global rigidity, 625–627
 Global simplicity, 652, 653
 Golden mean shift, 431
 G-operator system, 521, 522
 Graph-joining, 273
 Group endomorphisms, 587–588
 Group extensions of dynamical systems, 24
 Group measure space construction, 492, 507

H

Haar measure, 5
 Hahn–Banach theorem, 82
 Hajós–Minkowski theorem, 602
 Halmos topology, 537
 Halmos-von Neumann theorem, 395, 539
 Hamana boundary, 493, 520–522
 Hamiltonian function, 12
 Hamiltonian systems, 11–12
 billiard systems, 12–13
 KAM-systems and stably non-ergodic behavior, 13
 Hamilton's equations, 12, 534
 Hamming metric, 173
 Harrington-Kechris-Louveau dichotomy, 552
 Hausdorff a -measure, 64
 Hausdorff dimension, 348, 597, 600, 603, 606, 608, 638,
 642, 645, 647, 667, 669
 Hausdorff space, 497, 498, 506–508, 511, 513–515, 517,
 520, 522
 Height, 38
 Hellinger distance, 242
 Hénon diffeomorphisms, 22
 Hénon family of maps, 329
 Hénon-like maps, 646
 Hénon map, 656
 Herglotz's theorem, 67
 Heterodimensional homoclinic intersections, 654

Hidden Markov models (HMMs), 42
 Higher dimensional shift space, 427, 447–450
 Hilbert space, 36, 39, 52, 68, 492, 511, 521
 Hille-Yosida theorem, 479
 Hindman's theorem, 582–584
 Hoeffding's inequality, 43
 Hölder functions, 9
 Holonomy invariance, 640
 Homeomorphism, 494–498, 501–504, 506–508, 511, 512,
 514–517
 Homoclinic bifurcations, 669–670
 Homogeneous space, 597, 600, 601, 606
 Homomorphisms, 39
 Hopf Argument, 321–322
 Hopf's ergodic theorem, 95
 Horocycle flow, 15–16, 618
 Horseshoe maps, 17
 Host-Kra-Maass theorem, 417
 Host's theorem, 162
 Hurewicz pointwise Ergodic theorem, 236
 Hyperbolic (Anosov) dynamical system, 616
 Hyperbolic billiards and hard sphere gases, 327–328
 Hyperbolic flows, 586
 Hyperbolicity, 4, 14, 53, 633, 641–643, 646, 649, 650,
 653, 654, 656, 657
 Hyperbolic map, 589
 Hyperbolic matrix, 618
 Hyperbolic measure-preserving transformation, 4
 Hyperbolic set, 14
 Hyperbolic systems, 318, 641
 Hyperbolic toral automorphism, 586
 Hyperplane absolute winning (HAW) property, 608

I

Idempotent ultrafilter, 583
 Induced transformation, 24–25, 235
 Infinite direct product of finite types (IDPFT), 243
 Infinite divisibility, 217
 Infinite ergodic index, 253
 Infinite ergodic theory, 22–23
 Infinite iterated function system, 671
 Infinitely divisible, 217
 Infinite Markov shifts, 242
 Infinite measure preserving system, 65, 75
 Infinite partition, 10
 Infinite sequence, 375
 Infinite set, 582
 Information theory, 445
 Intermediate convergents, 10
 Intermittent maps, 21–22
 Internal disjointness, 303, 304
 Intersective polynomials, 70
 Intersymbol interference, 444
 Interval exchange maps, 6
 Interval exchange transformations (IETs), 38, 131–136,
 340–341, 351, 362
 joining properties, 360
 spectral measures of, 357–358

Interval maps, 383, 643–645, 649, 652, 654–657
 Invariant integrals, 462
 Invariant measures, 294, 374
 entropy of, 374–375
 Invariant probability measure, 635, 639, 644, 650–652
 Invariants of étale equivalence relations, 497
 Inverse limits, 27
 Invertible measure preserving system, 66, 70, 71
 Ising chain, 382
 Ising model, 370
 Isomorphism, 39, 493–498, 501, 502, 505, 507, 508, 510, 511, 516, 521, 522
 Iterated function systems, 669
 Iterated Kreiss condition, 479
 ITPFI transformation, 251–252

J

Jacobs-deLeeuw-Glicksberg decomposition, 477
 Jakobson one-dimensional theorem, 647
 Jakobson's method, 329
 Jewett–Krieger theorem, 402, 506
 Jewett–Krieger type realization, 494, 506
 Jiang–Su algebra, 516
 Jiang–Su stability, 516
 Joinings, 294, 298
 Joinings in ergodic theory
 and conjectures in number theory, 164
 definition, 149, 150
 disjointness, 153
 and factors, 155
 filtering problems, 160
 future research, 166
 and isomorphism, 154
 Markov intertwining, 156
 and multiple ergodic averages, 163
 Ornstein's and Krieger's theorems, 161
 and Rohlin's multifold mixing, 161, 163
 self-joining (*see* Self-joinings)
 set of, 152
 Julia set, 22, 213

K

Kac's formula, 235
 Kakutani equivalence, 211–212, 562–563, 573
 Kakutani–Rokhlin tower decomposition, 511, 517, 518
 Kakutani transformation, 241
 Kaplan–Yorke formula, 674
 Katok formulae, 650
 Katok fundamental class, 345
 K-automorphism, 258
 Keane conjecture, 347
 Keane's condition, 339
 Khintchine–Groshev Theorem, 600
 Khintchine's recurrence theorem, 75
 Khintchine's transference principle, 600
 Kingman's theorem, 91
 King's weak closure theorem, 278

Kneading invariants, 652, 653, 657
 Kochergin flows, 354
 Kolmogorov–Arnold–Moser (KAM) theory, 314
 Kolmogorov automorphism, 182
 Kolmogorov–Chaitin complexity, 371
 Kolmogorov extension theorem, 37
 Kolmogorov process (K-process), 181–182
 Kolmogorov–Sinai entropy, 205, 223
 Kolmogorov–Sinai theorem, 177
 Kontsevich–Zorich conjecture, 350
 Koopman mixing, 257
 Koopman operator, 36, 39, 221, 539
 Koopman representation
 Alexeyev's theorem, 115–116
 Alpern–Tikhonov topology, 123
 Banach problem, 112
 disjoint, 114
 entropy, 114
 ergodic dynamical system, 112
 Gaussian–Kronecker automorphism, 124
 Lamperti theorem, 113
 Markov operators, 113
 maximal spectral types, 114
 multiplicity (*see* Multiplicity function)
 pairwise independence property, 114
 spectral isomorphism, 112
 Thouvenot. Funny rank one, 123
 Koopman unitary operator
 associated with nonsingular Gaussian transformation, 269
 associated with nonsingular Poisson transformation, 268–269
 for nonsingular system, 267–268
 Kreiss bounded operators, 479
 Krengel class, 259–260
 Krengel entropy, 269
 Krengel–Pinsker factor, 270
 Krengel–Sucheston concept of mixing, 257
 Krickeberg mixing, 255
 Krieger embedding theorem, 442–443
 Krieger generator theorem, 177
 Krieger's theorem, 161, 246, 494, 505
 Kronecker factor, 48, 395
 Kronecker subset, 217
 Krylov–Bogoliubov theorem, 346, 537, 638

L

Lacunary sequence, 66
 Lagrange multiplier, 372
 Laplace–Beltrami operator, 585, 586
 Large deviations principles, 371
Lasota–Yorke inequality, 644
 Lebesgue density theorem, 639
 Lebesgue measure, 5–7, 9, 10, 18, 19, 22, 37, 45, 86, 316, 599, 605
 Lebesgue number, 185
 Lebesgue probability space, 27, 39, 42
 Lebesgue space, 23, 46

- Lebesgue spectrum, 116–117, 358
 - Ledrappier–Young formula, 646
 - Light mixing, 49
 - Lindenstrauss’s theorem, 102
 - Linear flow, 362
 - special representation of, 342
 - spectral measures of, 357–358
 - on torus, 335
 - on translation surfaces, 335–336
 - Linearization, 346
 - Liouvillean rotation number, 572
 - Liouville measure, 15
 - Liouville’s formula, 12
 - Liouville’s theorem, 534
 - Little shift equivalence problem, 440
 - Livshits’s criteria, 620
 - Local complexity, 651
 - Local entropy theory, 391
 - Local ergodicity, 642
 - Local ergodic theorem, 86
 - Local limit theorem, 649
 - Locally Hamiltonian flow, 333, 334, 336–337, 362–363
 - genericity notions for, 345
 - special representation of, 342–343
 - spectra of, 358–359
 - Local product structure, 16
 - Local rigidity, 623–625
 - Local stable manifolds, 14
 - Local unstable manifolds, 14
 - Local variational principle, 652
 - Logarithmic Furstenberg systems, 297, 306, 307
 - Logarithmic singularities, 342
 - Lower entropy factor theorem, 443–444, 452
 - Lusin’s separation theorem, 571
 - Lyapunov exponent, 19, 633, 636, 637, 639, 642, 643, 646, 650, 655, 674
 - Lyapunov spectra, 676
- M**
- Mackey Borel structure, 138
 - Maharam extension, 247–248
 - Mahler’s compactness criterion, 601
 - Mahler’s Criterion, 602
 - Mandelbrot set, 22
 - Mañé Stability Theorem, 641
 - Markov chain, 7, 40, 45, 47, 51, 446–447, 644
 - Markov-Feller operators, 463, 468
 - Markov intertwining, 149, 156
 - Markov maps, 382–383
 - Markov matrix, 179
 - Markov odometers, 240
 - Markov partition, 16, 430, 586, 587, 641
 - Markov process, 7, 179
 - Markov random fields, 214
 - Markov shift, 7, 37, 75, 179–180, 213, 633, 643, 653
 - Masur’s criterium, 348
 - Masur-Veech measure, 345
 - Matrix subshift, 448
 - Maximal analytic equivalence relation, 558
 - Maximal inequality, 82
 - Maximum entropy, 639, 640
 - measure, 640, 641, 643, 646, 652, 657
 - Maxwell-Boltzmann-Gibbs formula, 370
 - Möbius orthogonal, 1
 - Mean and pointwise ergodic theorems, 236–237
 - Mean dimension for a single homeomorphism, 511
 - Mean dimension for topological dynamical systems, 511
 - Mean ergodic theorem, 460–464, 579
 - Measurable distality, 411
 - Measure conjugacy, 550
 - Measure distal, 550
 - Measure isomorphism, 559
 - Measure of maximal entropy, 381
 - Measure-preserving dynamical system, 150
 - Measure-preserving homomorphism, 39
 - Measure-preserving systems, 4
 - Measure-preserving transformation, 36–40, 42–56
 - Measure rigidity, 627–628
 - Measures of maximal entropy, 427
 - Measure space, 184
 - Measure-theoretic dynamical system, 294
 - Measure-theoretic entropy, 36, 578, 640, 651
 - Measure-theoretic isomorphism, 39, 40
 - Mertens’ theorem, 586, 588
 - Meshalkin’s map, 178–179
 - Metric Diophantine approximation, 597, 598
 - Metric entropy, 8, 640
 - Metric number theory, 597, 599
 - Mildly mixing flow, 333
 - Mild mixing, 48, 49, 279, 353
 - Minimal action, 493–496, 498, 501, 503–505, 507, 511, 514, 515, 518–520
 - Minimal components, 339
 - Minimal homeomorphism, 529, 533, 550, 566
 - Minimality, 338
 - Minimal radon uniquely ergodic model, 239
 - Minimal self-joinings, 8, 149, 158, 353
 - Minkowski measure, 10
 - Mixing, 578, 586, 589–593
 - flow, 333
 - light, 49
 - mild-mixing, 48, 49
 - partially, 49
 - strong, 36, 39, 41, 45–51, 55
 - weak, 46–49, 52, 53, 55, 56
 - Möbius function, 420
 - aperiodicity, 304
 - logarithmic Furstenberg systems, 306, 307
 - multiplicativity, 303–304
 - short intervals, 304–306
 - Möbius orthogonality, 294, 295, 298–302
 - Möbius randomness principle, 164
 - Modulated ergodic theorem, 461–464, 476
 - Moment of order, 217
 - Morse function, 345
 - Morse shift, 432, 436
 - Morse-Smale diffeomorphism, 541, 563

- Multifold mixing, 162
- Multifractal analysis, 676, 677
- Multifractal decomposition, 676
- Multifractal spectrum, 676
- Multiple averages, 420
- Multiple recurrence, 69–71, 583
- Multiple weak mixing, 55
- Multiplicative orthogonality criterion (MOC), 303
- Multiplicativity
 - AOP property, 304
 - internal disjointness, 303, 304
- Multiplicity function
 - cocycle approach, 119–120
 - Gaussian and Poissonian automorphism, 120
 - maximal spectral type, 119
 - Rokhlin multiplicity problem, 121
- Multi-valued Hamiltonian flow, 336

- N**
- Natural extension, 27–28
- Newhouse phenomenon, 669
- Newhouse theorem, 657
- Newtonian mechanics, 4
- Newton's laws of motion, 4
- Nilmanifold, 592
- Nilrotation, 295
- Noisy channel theorem, 173–174
- Non-compact surfaces, 363
- Non-degenerate elliptic fixed point, 13
- Nonemptiness problem, 448
- Nonequilibrium steady states, 386
- Non-hyperbolic systems, 5
- Non-invertible transformations, 212
- Non-lacunary sequence, 67, 74
- Non-reversibility, 360
- Nonsingular Bernoulli shifts, 241–242
- Nonsingular Chacón map, 241
- Nonsingular dynamical system, 233, 234
- Nonsingular endomorphism, 243
- Nonsingular ergodic theory, 233
 - applications, 279–284
 - Bernoulli shifts, 241–242
 - Chacón map, 241
 - conservativity and ergodicity, 234–236
 - directional recurrence, 279
 - dynamical properties of Bernoulli and Markov shifts, 259–263
 - dynamical properties of IDPFT systems, 259
 - entropy, 269–272
 - ergodic decomposition, 237–238
 - future research, 284–285
 - Gaussian transformations, 243, 265
 - generator, 238
 - Glimm-Effros theorem, 238–239
 - IDPFT transformations, 243
 - joinings and factors, 272–276
 - K-automorphism, 258
 - Krickerberg mixing, 256
 - Markov odometers, 240
 - Markov shift, 242
 - mean and pointwise ergodic theorems, 236–237
 - minimal radon uniquely ergodic model, 239
 - multiple and polynomial recurrence, 258–259
 - natural extensions of nonsingular endomorphism, 243
 - nonsingular product odometers, 240
 - nonsingular restricted orbit equivalence, 272
 - nonsingular transformation, 234
 - orbit theory (*see* Orbit theory)
 - Parry's generalization, 271
 - Poisson suspension, 242–243, 263–264
 - rational ergodicity and rational weak mixing, 254–255
 - representations of ergodic flows, 239
 - smooth nonsingular transformation, 276–277
 - spectral theory, 265–269
 - topological groups, 244–246
 - tower transformation, 240
 - weak mixing, 252–254
- Nonsingular flows, 239
- Nonsingular joining, 273
- Nonsingular Markov shifts, 242, 262–263
- Nonsingular Poisson suspensions, 242–243
- Nonsingular product odometers, 240
- Nonsingular subadditive Ergodic theorem, 237
- Non-spectral approach, 481
- Non-uniform expansion/contraction, 645, 646
- Non-uniformly hyperbolic systems, 19
- Nonzero Lebesgue measure, 635
- Nonzero topological entropy, 635
- Normal number, 580, 581
- Novikov problem, 334, 335
- Nuclear dimension, 515
- Number theory, 598, 670

- O**
- 2-odometer, 6
- G-odometers, 508
- Off-diagonal self-joinings, 150, 157
- Operator ergodic theory, 460
 - general averaging methods, 474–476
 - mean ergodic theorem, 461–464
 - modulated ergodic theorem, 461–464
 - rates of convergence, 464–465
 - stability, 471–473
 - strong Cesàro convergence, 468–469
 - uniform ergodic theorem, 465–468
 - weak stability and mixing, 469–471
- Orbit complexity, 650
 - on the set of measures, 651
- Orbit-counting
 - exotic orbit growth, 588–589
 - and geodesics, 585–587
 - group endomorphisms, 587–588
 - Pólya–Carlson dichotomy, 589
- Orbit Dirichlet series, 578, 588
- Orbit equivalence of minimal actions of a finitely generated abelian group, 503

Orbit equivalence of minimal $\mathbb{Z}$ -actions on the Cantor set, 503
 Orbit theory, 246
 cocycles of dynamical systems, 250–251
 continuous orbit equivalence, 248–249
 full groups, 246–247
 ITPFI transformation, 251–252
 Maharam extension, 247–248
 normalizer of the full group, 249–250
 Ordered Bratteli diagram, 500
 Orientation preserving, 6
 isometries, 15
 Ornstein's isomorphism theorem, 206–207
 Ornstein's theorem, 161, 206
 Ornstein's theory, 56, 182
 Ornstein-Weiss tiling theorem, 511, 517
 Orthogonality of sequences, 295
 Orthonormal basis, 50
 Oseledec's theorem, 325
 Oseledec and Pesin theory, 674
 Oseledec's theorem, 642

P

Pairwise independence property (PID), 114
 Pairwise independently determined system (PID), 162
 Palis conjecture, 640
 PAPs, 229
 Parabolic systems, 671
 Parameter exclusion, 329
 Parry entropy, 270
 Partial hyperbolicity, 18, 324
 Partially hyperbolic dynamical systems, 18
 Anosov flows, time-one maps of, 18
 uniformly hyperbolic systems, compact group extensions of, 18
 Partially mixing, 49
 Pascal adic transformation, 56
 Periodic components, 339
 Periodic points, 381
 Perron-Frobenius theorem, 379–380
 Perron-Frobenius theory, 446
 Perron numbers, 452
 Pesin's theory of dimension-like characteristics, 668
 Pesin theory, 19, 325, 642, 656
 Phase space, 12, 35
 Phase space average, 638, 648
 Phase-structure grammars, 9
 Physical measure, 639, 656
 Piecewise C^2 expanding maps, 10
 Piecewise monotone maps, 652
 Pigeon hole principle, 236
 Pinsker algebra, 214
 Pinsker σ -ALGEBRA, 51
 Pinsker-algebra theorem, 182
 Pinsker factor, 181
 Pinsker field, 181
 Poincaré map, 12, 22

Poincaré recurrence, 64–65, 579, 583
 theorem, 390, 394
 Poincaré sequences, 394
 Point process, 217
 Pointwise dimension, 667
 Pointwise ergodic theorems, 88–90, 579
 Poisson entropy, 270–271
 Poisson factor, 226
 Poisson flow, 262
 Poisson point process, 1, 218–220
 Poisson processes, 219
 Poisson suspension, 220–223, 263–264, 281
 Polish group action, 530, 544, 551, 555, 557–558
 Polish space, 530–532, 545, 546, 549, 552, 553, 569, 570, 572
 Pólya–Carlson dichotomy, 589
 Polynomial deviations of ergodic averages, 350
 Polynomial rate (PR), 305
 Pontryagin duality, 540
 Positive definite sequence, 67
 Positive upper density, 582
 Power-bounded operator
 on Banach space, 471, 473
 ergodic, 464, 477
 on Hilbert space, 473
 on reflexive Banach space, 462
 Power series, 465
 Predecessor sets, 9
 Pressure functional, 668
 Prime number theorem, 585, 586
 Prime shift, 431
 Prime volume, 305
 Probabilistic limit theorems, 648
 Probability distribution, 374
 Probability measure, 8
 Probability space, 4, 39, 43, 44, 170–172
 Probability vector, 7, 171, 175, 177
 Product measure, 7
 Product transformation, 40
 Prouhet-Thue-Morse (PTM) sequence, 8, 432
 Proximal action, 520

Q

Qualitative theory, 533, 540
 Quantitative nondivergence method, 608
 Quantitative Poincaré recurrence, 64–65
 Quantitative weak mixing, 358
 Quasi-compact operator, 466
 Quasi-ergodic hypothesis, 36, 577, 579
 Quasi-genericity vs. logarithmic quasi-genericity, 295
 Quasi-hyperbolic automorphisms, 590

R

Radius of comparison, 512–514
 Radon-Nikodym derivative, 22, 244, 258
 Radon-Nikodym theorem, 85
 Ramsey theory, 420, 582–584

- Random dynamical systems, 24
- Random walks
 - boundaries of, 282
 - recurrence for, 281
- Rank function, 513
- Rank one attractors, 647
- Rank-one transformations, 38, 39, 45, 52, 55
- Rates of convergence, 464–465
- Rational ergodicity, 254–255
- Rational maps, 213
- Rational spectrum, 48
- Ratner property, 355
- Ratner's technique, 360
- Realization problem, 561
- Recurrence, 578, 579, 584, 590
 - combinatorics and number theory, 71–73
 - multiple, 69–71
 - quantitative Poincaré recurrence, 64–65
 - set of, 583
 - subsequence, 66–69
 - topological multiple, 582
- Recursive reduction, 536
- Reduced group C^* -algebra, 522
- Reduction property, co-analytic set, 572
- Regular expressions, 9
- Regular languages, 9
- Regular summability matrix, 474
- Relative entropy, 380–381
- Rényi inequality, 254
- Representations, realizations and genericity, 54–55
- Restricted orbit equivalence, 212
- Reversibility, 360
- Riemann hypothesis, 585
- Riemannian manifold, 14, 15, 18
- Riemannian metric, 15
- Riemann sphere, 22
- Riemann zeta function, 585
- Rigid flow, 333
- Rigidity, 615
 - differentiable, 620–623
 - future research, 628
 - global, 625–627
 - local, 623–625
 - measure, 627–628
 - sequence, 49, 279
- Rigid rotation, 5
- Rigid sequence, 128–130
- Ritt condition, 478
- Road-colorings, 441
- Road problem, 427, 442
- Rokhlin cocycle, 117, 121–122
- Rokhlin extensions of rotations, 306
- Rokhlin lemma, 514, 517
- Rokhlin property, 244
- Rokhlin's lemma, 26
- Rosendal theorem, 558
- Rotation number, 535, 536
- Roth's theorem, 75
- Roth-type condition, 350
- Rudolph's counterexample machine, 208
- Rudolph structure theorem, 207
- Ruelle's pressure formula, 669
- Ruelle transfer operator, 644
- Run-length limited shift, 428
- Ryzhikov's theorem, 163
- S**
- Saddle connection, 339
- Sarnak's conjecture, 165
 - Chowla conjecture, 299
 - Möbius orthogonality, 301
 - $(S_{\log})$ vs. $(C_{\log})$, 301
 - strategies, 302
 - strong MOMO property, 300
 - (S) vs. $(S_{\log})$, 302
 - zero entropy system, 298
- Second quantization operators, 221, 222
- Self-joinings, 150, 218
 - and commuting transformations, 157
 - minimal, 158
 - relative properties with respect to factor, 160
 - simple system, 159
- Semisimplicity, 160
- Sensitivity on initial conditions, 633, 637, 643
- Shannon-McMillan-Breiman theorem (SMB-theorem), 93, 180, 375
- Shannon's noisy channel theorem, 173–174
- Shape, 515, 517
- Shearing mechanism, 355, 360
- Shift equivalence, 428, 439, 440
 - strong, 428, 438, 439
- Shift map, 37
- Shift of finite type, 428, 433–435
- β -Shifts, 11
- Shift spaces, 428
 - higher dimensional, 447–450
 - invariant measures and entropy, 373–375
 - Ising chain, 382
 - Markov chains, finite alphabets, 381–382
 - measure of maximal entropy and periodic points, 381
 - run-length-limited, 444
 - shift of finite type and sofic shifts, 433–435
 - and sliding block codes, 431–433
- Short intervals, 303–306
- S_∞ -actions, 553–555
- Simplicity, 150
- Sinai billiards, 13
- Sinai–Ruelle–Bowen measure, 323, 371, 384, 634, 635, 639, 641–643, 646, 647, 655
- Singularities, 336
- Skew products, 24
 - group extensions of dynamical systems, 24
 - random dynamical systems, 24
- Slit torus, 349
- Smooth ergodic theory
 - absolute continuity, 322
 - Anosov diffeomorphism, 319
 - critical points and singularities, 327

- DA attractors, 320
 - definition, 313
 - dissipative partially hyperbolic diffeomorphisms, 325
 - distortion estimates, 320
 - examples of, 317–318
 - expanding maps, 319
 - Hopf Argument, 321–322
 - hyperbolic billiards and hard sphere gases, 327–328
 - hyperbolic blocks, 326
 - hyperbolic systems, 318
 - interval maps and parameter exclusion, 328
 - Lebesgue measure and local properties of volume, 316
 - near critical diffeomorphisms, 329
 - nonuniformly hyperbolic attractors, 326
 - partial hyperbolicity, 324
 - partially hyperbolic conservative diffeomorphism, 325
 - Sinai-Ruelle-Bowen measures, 323
 - Smooth flow, 333
 - cohomological equation for, 349–350
 - Smoothly isomorphic flows, 333
 - Smooth manifolds, 532, 540
 - Smooth measure-preserving transformations, 561–562
 - Smooth nonsingular transformation, 276–277
 - Smooth reparametrization, 338
 - Smooth transformations, 563
 - Sofic shift, 428, 433–435
 - Sofic systems, 9
 - Solenoids, 17–18
 - Space average, 579, 581
 - Spacers, 38
 - Spectral decomposition theorem, 17
 - Spectral disjointness, 359
 - Spectral measure, 218
 - Spectral properties, 36, 40
 - Spectral theorem, 112
 - Spectral theory, 82, 265–269
 - Banach problem, 118–119
 - Borel probability space, 116
 - 3-dimensional Heisenberg group, 136–138
 - Foias-Stratila theorem, 126–128
 - Gaussian dynamical systems, 125
 - Heisenberg odometer, 138
 - interval exchange transformations, 131–136
 - Lebesgue spectrum, 116–117
 - parabolic dynamical system, 130
 - perturbations of affine cocycles, 119–120
 - properties, 119
 - time-changes of nilpotent system, 131
 - time-changes of unipotent system, 130–131
 - weighted operators analysis, 116
 - zero entropy automorphism, 128
 - Stability, 653
 - Stable maps, 645
 - Standard Gaussian systems, 219
 - State splitting, 428, 438, 442
 - algorithm, 452
 - Stationary action, 282
 - Stationary Gaussian process, 219
 - Stationary process, 81, 218
 - Stationary stochastic process, 4, 81
 - Statistical learning theory, 42
 - Statistical properties, 648, 650
 - Statistical stability, 634, 654, 655
 - Stirling's formula, 172
 - Stochastic matrix, 7
 - Stochastic perturbations, 655
 - Stochastic stability, 634, 655
 - Stolz region, 478
 - Stone-Weierstrass theorem, 296
 - Stopping point, 338
 - Strict comparison, 513, 515–517, 519
 - Strict inequality, 650
 - Strong Cesàro convergence, 468–469
 - Strong mixing, 36, 39–41, 45–51, 55
 - Strong MOMO property, 307
 - Strong orbit equivalence, 494, 505, 506
 - Strong pinsker conjecture, 56
 - Structural stability, 14, 623, 634, 653, 654
 - Structure theory, 533
 - Sturmian systems, 8
 - Subsequence recurrence, 66–69
 - Subshift, 8–9, 295
 - finite type, 634, 641, 652
 - Subspace theorem, 590
 - Super-rigidity, 557
 - Suspension flows, 25
 - Switchable Ratner property, 355
 - Symbolic dynamics, 373–374, 428, 451–453, 637, 641–643, 652, 657
 - origins of, 429–431
 - Symbolic extension, 652
 - Symbolic modeling approach, 451
 - Symbolic shifts, 536–537
 - Symmetric logarithmic, 356–357
 - Synchronized automata, 304
 - Systems with discontinuities, 643, 644
 - Szemerédi's theorem, 69, 71, 72, 583, 584, 593
- T**
- Tail entropy, 651
 - Tail field, 181
 - Tent map, 635, 636, 638, 639, 642, 650, 652
 - Thermodynamic formalism, 453, 671
 - Thermodynamic formalism, for finite systems, 371–372
 - equilibrium distributions and Gibbs property, 372–373
 - finite lattice, 373
 - Time-change flow, 338
 - special representations, 344
 - Time-reparametrizations, 338
 - Time-zero partition, 176
 - Toeplitz systems, 9
 - Toms-Winter conjecture, 516, 517, 519
 - Topological complexity, 639, 640, 657
 - Topological conjugacy, 428, 433, 563–567
 - Topological distal, 548
 - Topological dynamical system, 295, 470, 493–495, 506–508, 511, 514

- Topological dynamics
 - chaos and complexity theory, 400–401
 - characteristic factors, 416–418
 - disjointness, 399–400
 - Ellis semigroup, 397–399
 - history, 390
 - Kronecker systems, 403–404
 - measurable mild mixing, 397
 - measurable mixing, 396
 - measurable weak mixing, 396
 - minimality and transitivity, 392–395
 - multiple ergodic averages, 418–420
 - problems, 391
 - recurrence, 391
 - structure theorems, 410–411
 - topological weak mixing, 396
 - weak disjointness, 400
 - Topological entropy, 4, 184–192, 428, 435, 446, 578, 586, 589, 637, 639, 640, 643, 650–652, 654, 657, 668
 - Topological genericity, 634, 641
 - Topologically transitive homeomorphism, 533
 - Topologically weak mixing, 352
 - Topological Markov chain, 371
 - Topological Markov shifts (TMS), 187
 - Topological mixing, 396
 - Topological multiple recurrence, 582
 - Topological orbit equivalence for actions, 492
 - Topological Pinsker factor, 404
 - Topological pressure, 668
 - Topological recurrence, 66, 73
 - Topology, 55
 - Toral automorphism
 - hyperbolic, 586, 587
 - quasi-hyperbolic, 587, 588
 - Totally non-symplectic (TNS) condition, 622
 - Tower, 514, 517, 518
 - map, 25
 - transformation, 240
 - Trachtman's approach, 442
 - Transfer function, 338
 - Transition matrix, 378
 - Transition probability, 461
 - Translation surface, 335–336
 - Trigonometric polynomial sequence, 476
 - Turbulence, 555
 - Twist map, 13
- U**
- Uniform approximation, 599
 - Uniform ergodic theorem, 465–468
 - Uniform hyperbolicity, 641, 642, 650, 653, 654, 656
 - Uniform Kreiss resolvent condition, 479
 - Uniformly expanding map, 641, 649, 652, 653, 655
 - Uniformly recurrent subgroup, 523
 - Uniform Rokhlin property, 514, 515
 - Unimodal maps, 20–21
 - Unimodular lattice, 601
 - Unique ergodicity, 45
 - Unique invariant measure, 651, 655
 - Uniquely ergodic model, 295
 - Unique trace property, 520, 522, 523
 - Unstable manifold, 14
 - Upper density, 67, 71, 72, 74
- V**
- Van der Corput's method, 68
 - Variational principle, 187, 189, 377, 650
 - Vertex shifts, 434
 - von Neumann algebras, 283, 492, 493
 - Von Neumann ergodic theorem, 68
 - von Neumann flows, 344, 353
 - rescaling of, 361
 - von Neumann mean Ergodic theorem, 44, 236
 - von Neumann's theorem, 82
- W**
- Wandering set, 22
 - Weak Bernoulli property, 40
 - Weak mixing, 39, 41, 46–49, 52, 53, 55, 56, 252–254
 - flow, 333
 - full measure of, 352
 - in linear flows and IETs, 351
 - logarithmic singularities, 357
 - quantitative, 358
 - Weak Pinsker conjecture, 56
 - Weak rational ergodicity, 254
 - Weak stability and mixing, 469–471
 - Weyl chamber flow, 619
 - Weyl rationally almost periodic (WRAP), 304
 - Weyl's criterion, 67
 - Wiener lemma, 111
 - Wiener–Wintner theorem, 96, 476, 477
- Y**
- Yomdin theory, 646
 - Young's dimension formula, 673
- Z**
- Zeeman's tolerance stability conjecture, 654
 - Zero-entropy, 180–181
 - continuous interval maps, 304
 - detection, 307
 - Gaussian system, 224
 - Zeta function, 428, 436, 437, 439, 577, 585–587, 589
 - $\mathbb{Z}^d$ -odometers, 509